

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:09:48 ; Search time 24.3 Seconds

(without alignments) 2393.655 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGTTTAEQALAAANSALA.....FTETIKVLDLTSIIORVE 1698

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7459	82.8	1693	1	P33424 hepatitis e
2	7448	82.6	1693	1	POLN_HEVPA
3	7433	82.5	1691	1	POLN_HEVME
4	7433	82.5	1693	1	POLN_HEVME
5	406.5	4.5	1648	1	RRPO_CGWS
6	374.5	4.2	2205	1	POLN_RUBVT
7	351.5	3.9	1612	1	RRPO_ORSVS
8	349	3.9	1844	1	POLN_TYVNC
9	346	3.8	1844	1	RRPO_ORSVS
10	339.5	3.6	1612	1	RRPO_ORSVS
11	322.5	3.6	1616	1	RRPO_TOML
12	321.5	3.6	1616	1	RRPO_TOML
13	320.5	3.6	1616	1	RRPO_TOML
14	320	3.6	1597	1	RRPO_CRMV
15	319.5	3.5	1616	1	RRPO_TMOB
16	316.5	3.5	1616	1	RRPO_TOMK2
17	313	3.5	1601	1	RRPO_TVCV
18	312.5	3.5	1612	1	RRPO_PMYV
19	312.5	3.5	1612	1	RRPO_PMYV
20	311	3.5	1748	1	POLN_ELV
21	310	3.4	1844	1	POLN_TYVNA
22	305.5	3.4	1776	1	POLN_OYV
23	301.5	3.3	1616	1	RRPO_TMYOM
24	299.5	3.3	1616	1	RRPO_TMYB
25	295.5	3.3	1616	1	RRPO_TMYV
26	293.5	3.3	1616	1	RRPO_TMYV
27	293	3.3	1874	1	POLN_KYVNA
28	291.5	3.2	1616	1	RRPO_TMYV
29	287.5	3.2	1839	1	POLN_EPMV
30	261.5	2.9	1609	1	RRPO_TMYV
31	254	2.8	1629	1	RRPO_TMYV
32	243.5	2.7	1707	1	RRPO_TMYV
33	243.5	2.7	1884	1	RRPO_ACTSP

34	242.5	2.7	1885	1	RRPO_ACTSP	P54891 apple chlor
35	218	2.4	2512	1	POLN_SINDV	P03317 sindbis vir
36	211	2.3	2514	1	POLN_SINDO	P27283 sindbis vir
37	207	2.3	2514	1	POLN_ONNGV	P13886 o'nyong-nyo
38	201.5	2.2	1718	1	RRPO_SHVX	P04575 shalilic vir
39	200	2.2	2431	1	POLN_SFV	P08411 semliki for
40	194	2.2	1967	1	RRPO_PMYV	P17965 potatato viru
41	194	2.2	2485	1	POLN_EEVV3	P36327 venezuelan
42	184.5	2.0	3175	1	RRPO_EAV	P18611 equine arte
43	177.5	2.0	2492	1	POLN_EEVT	P27282 venezuelan
44	176	2.0	1385	1	RRPO_P1AMV	P07518 plantago as
45	174.5	1.9	2492	1	POLN_EEVP	P36328 venezuelan

ALIGNMENTS

RESULT 1

POLN_HEVPA STANDARD: PRT: 1693 AA.

AC POLN_HEVPA

DT 01-FEB-1994 (Rel. 28, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE

(EC 2.7.7.48); HELICASE].

OS Hepatitis E virus (strain Pakistan) (HEV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage.

OX NCBI_Taxid=33774;

RN [1]

RP MEDLINE=92115700; PubMed=1731327;

RA Tsarev S.A., Emerson S.O., Reyes G.R., Tsareva T.S., Legters L.J.,

RA Malik I.A., Iqbal M., Purcell R.H.;

RT "Characterization of a prototype strain of hepatitis E virus.";

RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).

CC -! MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF

ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: M80581; AAA5725.1; -.

DR MEROPS: C41.001; -.

DR InterPro: IPR000606; -.

DR InterPro: IPR002588; -.

DR InterPro: IPR002589; -.

DR Pfam: PF01461; DUF27.1.

DR Pfam: PF01443; Viral_helicase1.1.

DR Pfam: PF01660; Methyltransf.1.

KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;

KW ATP-binding. 975

FT NP_BIND 975 ATP (POTENTIAL)

SO SEQUENCE 1693 AA; 185149 MW; 5A0F03FB1FB99E8 CRC64;

Query Match 82.8%; Score 7459; DB 1; Length 1693;

Best Local Similarity 81.8%; Pred. No. 0;

Matches 1401; Conservative 109; Mismatches 160; Indels 42; Gaps 7;

QY	1	PGTTTAEQALAAANSALANAVVPPFLSRVOTELLINIMOPROLVFPEVLMNPIOR	60
DB	10	PGTTTAEQALAAANSALANAVVPPFLSRVOTELLINIMOPROLVFPEVLMNPIOR	69
QY	61	VHNELEYCARAGCLEVGAHPSINDPNVLRHCFPLVGRDQVRYSAFTGPAAN	120
DB	70	VHNELEYCARAGCLEVGAHPSINDPNVLRHCFPLVGRDQVRYSAFTGPAAN	129

```

QY 121 CRRSALRGLPADRTYCFDFGSRCAFAETGVALYSLHDLMPADVAEAMAHGXTRLYAA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 CRRSALRGLPADRTYCFDFGSGCNPAETGIALYSLHDMSPDVAEAMRHGXTRLYAA 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 LHLPEVLLRPGYHTHTSYLLIHGDRAVYTYEGDTSAGYHHVSLIRATRTTKYGDH 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 LHLPEVLLRPGYHTHTSYLLIHGDRAVYTYEGDTSAGYHHVSNLRSMRTTKYGDH 249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 PLVIERRAIGCHFVLLTAAPESPMPYVYPRSTEVYASIGPGSPSPSPSACSTK 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 PLVIERRAIGCHFVLLTAAPESPMPYVYPRSTEVYASIGPGSPSPSPSACSTK 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 STEHAVVHTDRLMFGATLDQAFCSSRLMTYLRGISTKYVYGVALVANGMASDAL 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 STEHAVVHTDRLMFGATLDQAFCSSRLMTYLRGISTKYVYGVANGMASDAL 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 TAXTAATVITICHOYRLTOAISGMRLGVEHAKFTTRLYSMLPFKSGSDYIPGRLO 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 TAXTAATVITICHOYRLTOAISGMRLGVEHAKFTTRLYSMLPFKSGSDYIPGRLO 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 FYAOCRRMLSGHFLDPRVLVFEDESVPCRCRTFLKVAKGKFCFMRMLGDECTCLEPAE 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 FYAOCRRMLSGHFLDPRVLVFEDESVPCRCRTFLKVAKGKFCFMRMLGDECTCLEPAE 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 GLVGDHNDNEAYEGSEVDPAEPANLDVSGTYAVHGHOLEALYALVNPQDIARASRLT 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 GLVGDHNDNEAYEGSEVDPAEPANLDVSGTYAVHGHOLEALYALVNPQDIARASRLT 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 ATVELVASPDELCECRYLGNKTFRTYVVDGAHLBANGPEQVLSFSDASROSGASLSLT 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 ATVELVASPDELCECRYLGNKTFRTYVVDGAHLBANGPEQVLSFSDASROSGASLSLT 609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 YELTPAGLOVRISSNGLDCTATPPGAPASAPGEVAFAFCALRYNRNFTORHSLTGLW 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 YASASAGLETRYVAAGLDHRAVFAVSPRASAPGEVTAFCALRYNRNFTORHSLTGLW 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 661 LHPGGLIGTPPSPSGHIMESANPFCGEGTYLTWTS-----TSCFSSDFSPPE 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 670 FHPGGLIGTPPSPSGHIMESANPFCGEGTYLTWTS-----TSCFSSDFSPPE 724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 AAAPAMATGCLPHTSPVSDIWLPPSEEFQVDAVPA-PPAPDPA-GLGPVVLPP 767
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 725 PSLPSRAATP-----TP-----AAPLPPAPDPSPTLSAPRAGEBAP 761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 768 PPVYHKSIP-PPSRRRLTYTPDGAKVYAGSLFESDCDMLVNASVNGHRRGGLCHAF 826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 762 GATARAATHTQTRHRRLFTYTPDGSKVAFSGLFESTCTMLVNASVNDHRPGGLCHAF 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 827 YORPEAFYPTERTMREGLAATYLTTPRIIHAAPDYRVEDONPKRLERATRETSRGTA 886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 822 YORPEAFYPTERTMREGLAATYLTTPRIIHAAPDYRVEDONPKRLERATRETSRGTA 881
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 887 AYPLLGSGIYOVPSLSFDMERNHRPGDELTYLFEPAANPEAKPPPOVLTITEDART 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 882 AYPLLGSGIYOVPSLSFDMERNHRPGDELTYLFEPAANPEAKPPPOVLTITEDART 941
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 947 ANLALEIDAATEVGRACAGCTISPIGIYHOYTAGVPGSGKSRSTIOQGDVNVVPTRELR 1006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 942 ANLALEIDAATEVGRACAGCTISPIGIYHOYTAGVPGSGKSRSTIOQGDVNVVPTRELR 1001
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1007 NSNRBRGAFTPHRTAARVITIGRRVVIDEAPSLPRLLLHMHORASSVHLLGDPNQIPAI 1066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1002 NSNRBRGAFTPHRTAARVITIGRRVVIDEAPSLPRLLLHMHORASSVHLLGDPNQIPAI 1061
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1067 DEFHAGIVPAIRPELATSMWXYTHRCRADVCELIRAGAYKIOGTSVLSLFENEPAG 1126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1062 DEFHAGIVPAIRPELATSMWXYTHRCRADVCELIRAGAYKIOGTSVLSLFENEPAG 1121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1127 OKLVXTQAAANPAGATVVEAGATETETIITADARGLIOSSRAHAIALTRHTEKC 1186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1122 OKLVXTQAAANPAGATVVEAGATETETIITADARGLIOSSRAHAIALTRHTEKC 1181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 VIIDAPGLLEVGISDIYVNNFLVAGSEVGHRSPIVRGNPDONLGTLOAFPSCQISA 1246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1182 VIIDAPGLLEVGISDIYVNNFLVAGSEIGHORSVPIRGNPDANVDTLAAFPSCQISA 1241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1247 YHOLAEELGHRPAVAANVLPCCPELBOGLLYMPELVSDSVYFEELTDIVHCHMAAPSO 1306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1242 YHOLAEELGHRPAVAANVLPCCPELBOGLLYMPELVSDSVYFEELTDIVHCHMAAPSO 1301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1307 RKAVALSTLVGRYGRRTKLYEAASDVRESLARIPRTIGPVQATTCELYELVEAMVERGOD 1366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1302 RKAVALSTLVGRYGRRTKLYEAASDVRESLARIPRTIGPVQATTCELYELVEAMVERGOD 1361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1367 GSAVLELDLCNRDVSRTTFQKXCNKFTTGETTAHGVGGISAMSTFCALRPPMRRAI 1426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1362 GSAVLELDLCNRDVSRTTFQKXCNKFTTGETTAHGVGGISAMSTFCALRPPMRRAI 1421
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1427 EKELIALLPIEFGDYEESVFAAASGAGSCVFNDESEFSTONNFSIGLECYME 1486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1422 EKALIALLPQGVYFGADPDTVFSAAYAAKASVFNDESEFSTONNFSIGLECYME 1481
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1487 ECGMPOMLIRLYHLVRSAMILLQAPKESLKGFWKHSGEPGLTMNTVNMMAITAHCEYFR 1546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1482 ECGMPOMLIRLYHLVRSAMILLQAPKESLKGFWKHSGEPGLTMNTVNMMAITAHCEYFR 1541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1547 DFRVAARFGDSDVYLCSDYRQSRNAALIAAGCGIKLVDRPRTGLYAGVVAAPGLTLPD 1606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1542 DLOVAAERGDSDVYLCSEYRQSPGAVALIAGCGIKLVDRPRTGLYAGVVAAPGLTLPD 1601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1607 VVRPAGRLSEKNMGPGPERAEQLRLAVCDFLRLGTLYNNAQVCVDVYASVYGVSPGLVNL 1666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1602 VVRPAGRLSEKNMGPGPERAEQLRLAVCDFLRLGTLYNNAQVCVDVYASVYGVSPGLVNL 1661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1667 GMLQTIADGKAHFTETIKPVLDLTNSIIORVE 1698
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1662 GMLQTIADGKAHFTESYKPVLDLTNSIIORVE 1693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 2
 ID POLN_HEVBU STANDARD; PRT; 1693 AA.
 AC P29324;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
 (EC 2.7.7.48); HELICASE].
 OS Hepatitis E virus (strain Burma) (HEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 ON NCBI_TaxID=31767;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92024067; PubMed=1926770;
 RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
 RA Fry K.E., Reyes G.R.;
 RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
 RT full-length viral genome."
 RL Virology 185:120-131(1991).
 RL MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
 ENTERICALLY TRANSMITTED NON-A; NON-B HEPATITIS (ET-NANBH).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).
 CC EMBL; M73218; AAA45734.1;
 CC DR PIR: A40778; MNMHE.
 CC MEROPS: C41.001; -;
 CC InterPro: IPR000606; -;
 CC InterPro: IPR002588; -;
 CC InterPro: IPR002589; -;

DR pfam: PF01661: DUF27.1.
 DR pfam: PF01443: viral_helicase1.1.
 DR pfam: PF01660: methyltransferase.1.
 KW Polyprotein; transferase; RNA-directed RNA polymerase; Helicase;
 FT ATP-binding. 975
 FT NP_BIND 982 ATP (POTENTIAL).
 SQ SEQUENCE 1693 AA; 185191 MW; 2F355E46E9ED219B CRC64;

Query Match 82.6%; Score 7448; DB 1: Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 196; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 1 PGGTTTAEQAALAAANSAANAVVVPFLSRVOTELLINLMOPROLYFREVLMNPIOR 60
 DB TT
 DB 10 PGGTTTAEQAALAAANSAANAVVVPFLSHOELLINLMOPROLYFREVLMNPIOR 69
 QY 61 VHNLEEOYCRARAGCLEVGAPHSRINDPNVLRHRCFLRPVGRDVRWYSAFTGPAAN 120
 DB TT
 DB 70 VHNLEELCYCRARSGCLEIGAPHSRINDPNVLRHRCFLRPVGRDVRWYSAFTGPAAN 129
 QY 121 CRRSALRGLPPADRTYCFDFGSCFAETVALYSLHDLMPADVAEAMARHGXTRLYAA 180
 DB TT
 DB 130 CRRSALRGLPPADRTYCFDFGSCNPAETGALYSLHDMSPSDVAEAMRHGXTRLYAA 189
 QY 181 LHLPPVLLPPTGYTHTSTYLLHDSRAVVTYEGDTSAGYNHDVSLTRAMRTKTVGDH 240
 DB TT
 DB 190 LHLPPVLLPPTGYTHTSTYLLHDSRAVVTYEGDTSAGYNHDVSLTRAMRTKTVGDH 249
 QY 241 PLVIEVRAIGCFHVLTLTAAPESPMPVYPRSTEVYRSIFGCGSPSLPSCSTK 300
 DB TT
 DB 250 PLVIEVRAIGCFHVLTLTAAPESPMPVYPRSTEVYRSIFGCGSPSLPSCSTK 309
 QY 301 STPHAVVHIMDLMLFGATLDDQAFCCSLMTYLRGISYKTVGALVANEGMNASEDAL 360
 DB TT
 DB 310 STPHAVVHIMDLMLFGATLDDQAFCCSLMTYLRGISYKTVGALVANEGMNASEDAL 369
 QY 361 TXXITAAVLTICHOYLRTOALSCKMRRLGVEHAOKFTRRLYSWLEKSGROVTPROQ 420
 DB TT
 DB 370 TXXITAAVLTICHOYLRTOALSCKMRRLGVEHAOKFTRRLYSWLEKSGROVTPROQ 429
 QY 421 FYAOCRRMLTSGFHLDPRLVLPDESVPKRCRTFLKRVAGKFCFCFMMWLGCECTFLEPA 480
 DB TT
 DB 430 FYAOCRRMLTSGFHLDPRLVLPDESVPKRCRTFLKRVAGKFCFCFMMWLGCECTFLEPA 489
 QY 481 GLVGDHGDNEAESEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPDIIAARSLRT 540
 DB TT
 DB 490 GLVGDHGDNEAESEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPDIIAARSLRT 549
 QY 541 ATVELVADRLCETVLTGKTRFTVVDGAHLEANGRPQVYLSFASRQSGMAGSHSLT 600
 DB TT
 DB 550 ATVELVADRLCETVLTGKTRFTVVDGAHLEANGRPQVYLSFASRQSGMAGSHSLT 609
 QY 601 YELTPAGLOVRLSSNGLDCTATFPFGAPSAAGEVAACFALRYNRFORSLTGLW 660
 DB TT
 DB 610 YASASAGLEVRVAAGLDRAVFAPEVSPRSAGCEVATACALYRRNREQRSLGILNM 669
 QY 661 LHPGLIGTFPPSPGHIMESANPCGEGTLTYRTWS-TSGESSDPS-PEAARA 714
 DB TT
 DB 670 LHPGLIGTFPPSPGHIMESANPCGEGTLTYRTWS-EVDAVSPARPDLGFMSEPSIPS 729
 QY 715 MATPGLPSTPVSIDVWLPPPESEFOYDAVY-PPAPPAFL-GRVYLTPPPPPVH 772
 DB TT
 DB 730 MATPGLPSTPVSIDVWLPPPESEFOYDAVY-PPAPPAFL-GRVYLTPPPPPVH 776
 QY 773 KPSIP-PPSRNRLTYTDPGAVYASLFESECDMLVNASNPGHPPGGGLGHAFTYQRP 831
 DB TT
 DB 767 KPSIP-PPSRNRLTYTDPGAVYASLFESECDMLVNASNPGHPPGGGLGHAFTYQRP 836
 QY 832 EAFYPTFETIRBGLAATYTLPRITIAVADYVEONPKLEAAYETCSRGSTAYPPL 891
 DB TT
 DB 827 EAFYPTFETIRBGLAATYTLPRITIAVADYVEONPKLEAAYETCSRGSTAYPPL 886

QY 892 GSGIYQVPSLSTDAWERNHRCDELYLTERPAANFEAKPKAPVLTITTEDTARTANL 951
 DB TT
 DB 887 GSGIYQVPSLSTDAWERNHRCDELYLTERPAANFEAKPKAPVLTITTEDTARTANL 946
 QY 952 EIDAEVGRACAGCTISGIYHYOFTAGVPGSGKRSIQGQDVVVVPTTELRSNMR 1011
 DB TT
 DB 947 EIDAEVGRACAGCTISGIYHYOFTAGVPGSGKRSIQGQDVVVVPTTELRSNMR 1006
 QY 1012 RGFPAFTPHTAARVTIGRRVYIDEAPSLRPHLLLMORASSVHLGDPNOIPAIDEHA 1071
 DB TT
 DB 1007 RGFPAFTPHTAARVTIGRRVYIDEAPSLRPHLLLMORASSVHLGDPNOIPAIDEHA 1066
 QY 1072 GLVPAIRPELAPTSVMXVYHRCADYCELLRGAYPKIOTTSRLSLFNENPAIGOKLX 1131
 DB TT
 DB 1067 GLVPAIRPELAPTSVMXVYHRCADYCELLRGAYPKIOTTSRLSLFNENPAIGOKLX 1126
 QY 1132 TOAKANFGAIVTVEAOGATFETTLTADARGLIOSSRAHAYALRHTREKCYLDA 1191
 DB TT
 DB 1127 TOAKANFGAIVTVEAOGATFETTLTADARGLIOSSRAHAYALRHTREKCYLDA 1186
 QY 1192 PGLREVGISDVIVNNEFLAGEVGHXRPVYPRGNPDONLTLQAPPSCOISATYHOLA 1251
 DB TT
 DB 1187 PGLREVGISDVIVNNEFLAGEVGHXRPVYPRGNPDONLTLQAPPSCOISATYHOLA 1246
 QY 1252 EELGHRPAAVAPLPPCPLEOGLLYMPOELTVSDSVLVELTDIYHGMARSPORKAVL 1311
 DB TT
 DB 1247 EELGHRPAAVAPLPPCPLEOGLLYMPOELTVSDSVLVELTDIYHGMARSPORKAVL 1306
 QY 1312 STLVRGRTKLYLEAASDVRESLARFTPTGPOATTCCELYLEVAWEKQDGSAYL 1371
 DB TT
 DB 1307 STLVRGRTKLYLEAASDVRESLARFTPTGPOATTCCELYLEVAWEKQDGSAYL 1366
 QY 1372 ELDLCNRDYSRTFFPKXCKFTTGETIAHGKYGQGISAMSKTFCLFPGWFAITEKEL 1431
 DB TT
 DB 1367 ELDLCNRDYSRTFFPKXCKFTTGETIAHGKYGQGISAMSKTFCLFPGWFAITEKEL 1426
 QY 1432 ALPPIVIFGDAVEESVFAAAGSAGSCWFEVDFEPTSTONFSLGECVMEECGMP 1491
 DB TT
 DB 1427 ALPPIVIFGDAVEESVFAAAGSAGSCWFEVDFEPTSTONFSLGECVMEECGMP 1486
 QY 1492 QMLIRLYHLVRSAMILQAPRESLKGFKHSGEGTLNMTVMNMAIIACVFEFRFVA 1551
 DB TT
 DB 1487 QMLIRLYHLVRSAMILQAPRESLKGFKHSGEGTLNMTVMNMAIIACVFEFRFVA 1546
 QY 1552 AFKGDSDVLCSDYROSRNAALIIACGKLKVDYPRIGLYAGVVAAPGLTLPDYVRR 1611
 DB TT
 DB 1547 AFKGDSDVLCSDYROSRNAALIIACGKLKVDYPRIGLYAGVVAAPGLTLPDYVRR 1606
 QY 1612 GRLEKMGPGPERAOLRLAVCDPLRGLTNVQVCVDVSVRYGSPGLVYHNLIGMLQT 1671
 DB TT
 DB 1607 GRLEKMGPGPERAOLRLAVCDPLRGLTNVQVCVDVSVRYGSPGLVYHNLIGMLQT 1666
 QY 1672 IADKRAHFTETIKPVLDTNLSIIQRYE 1698
 DB TT
 DB 1667 IADKRAHFTETIKPVLDTNLSIIQRYE 1693

RESULT 3
 POLN_HEVME
 ID POLN_HEVME STANDARD; PRT; 1691 AA.
 AC 003495;
 DT 01-OCT-1993 (rel. 27, Created)
 DT 01-OCT-1993 (rel. 27, Last sequence update)
 DT 15-DEC-1998 (rel. 37, Last annotation update)
 DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
 (EC 2.7.7.48); HELICASE].
 OS Hepatitis E virus (strain Mexico) (HEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_TaxID=31768;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93079857; PubMed=1448913;
 Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,

RA Bradley D.W., Tam A.W., Reyes G.R.;
 RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
 RL E virus (HEV)."
 RN Virology 191:550-558(1992).
 RP
 RX SEQUENCE OF 965-1691 FROM N.A.
 RA MEDLINE-92271462; PubMed-158964;
 RA Ery K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
 RA Platack M., Feldman R.A., Yun K.Y., Purdy M.A., McCaustland K.A.,
 RA Bradley D.W., Reyes G.R.;
 RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
 RT region encoding consensus motifs for an RNA-dependent RNA polymerase
 RT and an ATP/GTP binding site."
 RL Virus Genes 6:173-185(1992).
 CC -I- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
 CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M74506; AAA45730.1; -
 DR PIR; A44212; A44212.
 DR MEROPS; C41.001; -
 DR InterPro: IPR000606; -
 DR InterPro: IPR002588; -
 DR InterPro: IPR002589; -
 DR Pfam; PF01661; DUF27; 1.
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR Pfam; PF01660; Vmethyltransf; 1.
 KM Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
 KM ATP-binding; 973 980 ATP (POTENTIAL).
 FT NP_BIND 1691 AA; 185224 MW; DB5F0B2C913F871B CRC64;
 SQ SEQUENCE

Query Match 82.5%; Score 7433; DB 1; Length 1691;
 Best local similarity 81.0%; Pred. No. 0;
 Matches 1384; Conservative 121; Mismatches 166; Indels 38; Gaps 5;

QY 1 PGITTAEQAAALAAANSAANAVVYRPFLSRVQTEILINLMQPROLYRPREVLMNHPQR 60
 DB 10 PGITTAEQAAALAAANSAANAVVYRPFLSRVQTEILINLMQPROLYRPREVLMNHPQR 69
 QY 61 VINHELEQYCARAGRCLEVAHRSINDNPNVLRHCLRVGVGVQVQWYATPRGPAN 120
 DB 70 VINHELEQYCARAGRCLEVAHRSINDNPNVLRHCLRVGVGVQVQWYATPRGPAN 129
 QY 121 CRRSALGLPRADRTYCGDFGSCFAFAETVALYSLHDLMPADVAEAMRHGRTLYAA 180
 DB 130 CRRSALGLPRADRTYCGDFGSCFAFAETVALYSLHDLMPADVAEAMRHGRTLYAA 189
 QY 181 LHPREVLPRGYHTTSYLLIHGDRAVVTYEGDTSAGYNHDSILRMTTRTKIVGDH 240
 DB 190 LHPREVLPRGYHTTSYLLIHGDRAVVTYEGDTSAGYNHDSILRMTTRTKIVGDH 249
 QY 241 PLVTERVATGCHVLLTAAPEPSPMYVPRSTEVYVSTFGPGSPSLFPGACSTK 300
 DB 250 PLVTERVATGCHVLLTAAPEPSPMYVPRSTEVYVSTFGPGSPSLFPGACSTK 309
 QY 301 STEHAVVHIWDRMLFGATLDDAFCSSRLMTYLRGISYKVTYGAALVANGMNASDAL 360
 DB 310 STEHAVVHIWDRMLFGATLDDAFCSSRLMTYLRGISYKVTYGAALVANGMNASDAL 369
 QY 361 TATTAAYLTICHOVRYLRTQAISKGRRLGVEHAOKFTIRLYSMLFEKSGRDYIPGRLO 420
 DB 370 TATTAAYLTICHOVRYLRTQAISKGRRLGVEHAOKFTIRLYSMLFEKSGRDYIPGRLO 429
 QY 421 FYACGRMLSGFLHDLPRVLVFDSEVPCRCRTFLKVAAGKCCFMRLMGQECTCFLEPAE 480

DB 430 FYACGRMLSGFLHDLPRVLVFDSEVPCRCRTFLKVAAGKCCFMRLMGQECTCFLEPAE 489
 QY 481 GLVGDHGDNEAREGSEVDPAEPALDVSQGYAVHGHOLEALRYLANVPODIAARASLT 540
 DB 490 GLVGDHGDNEAREGSEVDPAEPALDVSQGYAVHGHOLEALRYLANVPODIAARASLT 549
 QY 541 ATVELVASPDRLECRVYLGKNTFRTTVVVGALHEANGPEQVLSFDSAQSGAGSHSLT 600
 DB 550 ATVELVASPDRLECRVYLGKNTFRTTVVVGALHEANGPEQVLSFDSAQSGAGSHSLT 609
 QY 601 YELPAGIQVATISSNGLDCTATFPFGCAPSAAPGEVAAFGALRYNFTORHSLTGLM 660
 DB 610 YELPAGIQVATISSNGLDCTATFPFGCAPSAAPGEVAAFGALRYNFTORHSLTGLM 669
 QY 661 LHPGLGIFPPFSGHGWESANPFCGGTLYTRTWST-----SGFSSDFSPPE 709
 DB 670 LHPGLGIFPPFSGHGWESANPFCGGTLYTRTWST-----SGFSSDFSPPE 728
 QY 710 AAAPAMATPOLPHSTPEVSDIWLPPSEEFQVDAAPVPAPAPAGLPGVVLTPPPP 769
 DB 729 GGPATATGPAVGSDDSDPD-PLPDVTDGSRSGAR-PAGPMPNGVP----- 774
 QY 770 PVKRPSTLPPSRNRLTYTPDGAKVYAGSLPESDCMLVNASNGRHRPGGLCHAFYQR 829
 DB 775 -----QRRLLHTYPDGAKIYVGSIFSECTWLVNASNAGHRPGGLCHAFYQR 822
 QY 830 FPEAFYPEFIMREGIAAYTLTPRPIIHAVAPDVVRVONPRLAEAYRENCSSRRGTAYP 889
 DB 823 YPDSFDATKFEVMDGLAYTLTPRPIIHAVAPDVVRVONPRLAEAYRENCSSRRGTAYP 882
 QY 890 LLGSGIYQVPSLSEFDMERNHRBGDELXYLTPAPANFEAKPAQVYLITTEDARTANL 949
 DB 883 LLGAGIYQVPSLSEFDMERNHRBGDELXYLTPAPANFEAKPAQVYLITTEDARTANL 942
 QY 950 ALEIDATEVRACAGCTISGCIYHYOFTAGVPSSGSKRSIQQGDVYVYPTTELRSW 1009
 DB 943 ALEIDSGSEVGRACGKVEGVYQFTAGVPSSGSKRSIQQGDVYVYPTTELRSW 1002
 QY 1010 RRRGFAFTPTTAARVITGRVVIDEAPSLPRLHLLHMOASSVHLLGDNPQIPADFE 1069
 DB 1003 RRRGFAFTPTTAARVITGRVVIDEAPSLPRLHLLHMOASSVHLLGDNPQIPADFE 1062
 QY 1070 HAGLVPALRPELAPTSMWXYVTHRCPADVCELIRGAPKRIQTSYVLSLFWNEBAGOKL 1129
 DB 1063 HTGLIPALRPELAPTSMWXYVTHRCPADVCELIRGAPKRIQTSYVLSLFWNEBAGOKL 1122
 QY 1130 VXTQAARAANGCATVHEAGCATFETTTIATAGLGIOSRHAAYALVLRHREKCYIL 1189
 DB 1123 VXTQAARAANGCATVHEAGCATFETTTIATAGLGIOSRHAAYALVLRHREKCYIL 1182
 QY 1190 DAPGLREVGISDYIVNFFLAGGEVGHXRPVYIPRGNDPQNLCTLOAFPSCQISAYHQ 1249
 DB 1183 DSPGLREVGISDYIVNFFLAGGEVGHXRPVYIPRGNDPQNLCTLOAFPSCQISAYHQ 1242
 QY 1250 LAELGHRPAPVAAVLPCEPLEBGLLYMPOELTVSDSVLEFLTDIVHCMAAPSORKA 1309
 DB 1243 LAELGHRPAPVAAVLPCEPLEBGLLYMPOELTVSDSVLEFLTDIVHCMAAPSORKA 1302
 QY 1310 VLSTLVGRYGRRTLYEAAHSDVRESLARFTPTGPAQVATCELYELVEAVNEVGGGSA 1369
 DB 1303 VLSTLVGRYGRRTLYEAAHSDVRESLARFTPTGPAQVATCELYELVEAVNEVGGGSA 1362
 QY 1370 VLELDLGNRDSRTTFQCKXCKNFTTETIANGVVGOGISAMSKTFPCALRGPMRALEKE 1429
 DB 1363 VLELDLGNRDSRTTFQCKXCKNFTTETIANGVVGOGISAMSKTFPCALRGPMRALEKE 1422
 QY 1430 ILALPPIVIFYGDAYEESVFAAASGAGSCWPFENDFSEPDSTONNPSLGLCYVMEBCG 1489
 DB 1423 ILALPPIVIFYGDAYEESVFAAASGAGSCWPFENDFSEPDSTONNPSLGLCYVMEBCG 1482
 QY 1490 MPQWLIRLHYVRSAMTLOAPKESLKGFWKKHSEPEPTLLMNTYWNNAIIAHCEPFEDFR 1549

Db 1483 MEOWLRLYHAVRSAMIIQAPKESLRGFKHSGEPGSLMNTVMNMAIIAHCEFRDQ 1542

QY 1550 VAAFGDDSVVLCSDYRGRNAAALAGCGLLKDYRIGLYAGVYAPGCTLPDVR 1609

Db 1543 VAAFGDDSVVLCSEYRSGRPGAGSLIAGCGLLKADFRDIGLYAGVYAPGALPDVR 1602

QY 1610 FAGRISEKMGWGPAREQRLIAVCDLGLTNVAQVADVYRVSRYGVSPLVHNLIGML 1669

Db 1603 FAGRISEKMGWGPAREQRLIAVCDLGLTNVAQVADVYRVSRYGVSPLVHNLIGML 1662

QY 1670 QTADGKAHFTETIKPVDLTNSIQRYE 1698

Db 1663 QTIGDKAHFTESVAKPIDLTHSIMHRE 1691

RESULT 4

AC POLN_HEVMY STANDARD: PRT: 1693 AA.

AC 004610:

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48); HELICASE].

OS Hepatitis E virus (strain Myanmar) (HEV).

OC Viruses: ssRNA positive-strand viruses, no DNA stage.

OX NCBI_TaxID=31769;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93227573; PubMed=8470371;

RA Aye T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M., Rikhisia T., Wain K.;

RT "Sequence and gene structure of the hepatitis E virus isolated from Myanmar";

RL Virus genes 7:95-109(1993).

CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL: D10330; BAA01172.1; -

DR MEROPS: CA1.001; -

DR InterPro: IPR000606; -

DR InterPro: IPR002588; -

DR InterPro: IPR002589; -

DR Pfam: PF01661; DUF27.1;

DR Pfam: PF01443; Viral_helicase1.1;

DR Pfam: PF01660; Vmethyltransf.1;

KM Polypeptide: Transferase; RNA-directed RNA polymerase; Helicase;

KM ATP-binding. 975 982 ATP (POTENTIAL).

FT NP_BIND 1693 AA; 185215 MW; AABAC9140A/E21EA CRC64;

SO SEQUENCE

Query Match 82.5%; Score 7433; DB 1: Length 1693;

Best local similarity 81.1%; Pred. No. 0;

Matches 1394; Conservative 109; Mismatches 159; Indels 56; Gaps 6;

QY 2 GITTAIEQAALAAANSALANAVVVRPFLSRVQTEILINIMQFROLYFREVLMNHPIQRY 61

Db 11 GITTAIEQAALAAANSALANAVVVRPFLSRVQTEILINIMQFROLYFREVLMNHPIQRY 70

QY 62 IHNELEQYCRAGRCLEAGHPRSRINDPNVLRHCFLEPVRGVQRYTSAPTRGPANC 121

Db 71 IHNELEQYCRAGRCLEAGHPRSRINDPNVLRHCFLEPVRGVQRYTSAPTRGPANC 130

QY 122 RRSALRGLEPPADRYTCFQDSFRCFAFETGVALYSLHDLMPADVNAEMARHGXTRLYAL 181

Db 131 RRSALRGLEPPADRYTCFQDSFRCFAFETGVALYSLHDLMPADVNAEMARHGXTRLYAL 190

QY 182 HLEPEVLLPQGYHTTSYLLIHDGDAVYVEGDTSAGYNHDVSLIRANIRTKIYGDHP 241

Db 191 HLEPEVLLPQGYHTTSYLLIHDGDAVYVEGDTSAGYNHDVSLIRANIRTKIYGDHP 250

QY 242 LVIERVRAIGCFVLLTAAPRSPMPYVPRSTRTEVYRSIFPGGSGSLFSSACSTRS 301

Db 251 LVIERVRAIGCFVLLTAAPRSPMPYVPRSTRTEVYRSIFPGGSGSLFSSACSTRS 310

QY 302 TFHAVPVHIDRLMLFGATLDQAFCCSRMLMTYNGISKVNVGALVAEGNNSADAT 361

Db 311 TFHAVPVHIDRLMLFGATLDQAFCCSRMLMTYNGISKVNVGALVAEGNNSADAT 370

QY 362 AXITAAVLTICQRYLRTQAIKGMRLGVEHAQKFTPLYSMLFEKSGRDYIPGROLQF 421

Db 371 AXITAAVLTICQRYLRTQAIKGMRLGVEHAQKFTPLYSMLFEKSGRDYIPGROLQF 430

QY 422 YACRRMLTSAGHLDPRVLYFDESVPCRCRTFLKVAAGKCCFPMKLGOCCTCFLEPAG 481

Db 431 YACRRMLTSAGHLDPRVLYFDESVPCRCRTFLKVAAGKCCFPMKLGOCCTCFLEPAG 490

QY 482 LVGDGHNDNEAYEGSEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPQDIARAASRLTA 541

Db 491 LVGDGHNDNEAYEGSEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPQDIARAASRLTA 550

QY 542 TVELVASPDLCEKRYLGNKRTERTVVDGAHLHANGPEQYVLSFPAASROKSGAGSHLYT 601

Db 551 TVELVASPDLCEKRYLGNKRTERTVVDGAHLHANGPEQYVLSFPAASROKSGAGSHLYT 610

QY 602 ELTPRGLOVIRISNGIDCATFPFGAPSAAPGVAAPFCALYRNRFORHSTIGGLML 661

Db 611 ELTPRGLOVIRISNGIDCATFPFGAPSAAPGVAAPFCALYRNRFORHSTIGGLML 670

QY 662 HPEGLGIPPPSPGHIMESANPFCGEGTLTWTMS-TGSESSDESP-----PEAAAPAM 715

Db 671 HPEGLGIPPPSPGHIMESANPFCGEGTLTWTMS-TGSESSDESP-----PEAAAPAM 730

QY 716 AATPGLPHSTPVSDIWLVPPESEFQVDAAPVPAAPDAGLPGVYLTPPPPPYHKPS 775

Db 731 AATPGLPHSTPVSDIWLVPPESEFQVDAAPVPAAPDAGLPGVYLTPPPPPYHKPS 785

QY 776 IPPP-----SNRRLLYTPPDGAKYVAGSLFESDCMLVNASNPGIRPGG 820

Db 786 IPPP-----SNRRLLYTPPDGAKYVAGSLFESDCMLVNASNPGIRPGG 835

QY 821 GLCHAFYQRFPAFYPTFTFIREGLAAYTLTPRPIIHAVAPDYRVQNPKRLEAAYRETC 880

Db 831 GLCHAFYQRFPAFYPTFTFIREGLAAYTLTPRPIIHAVAPDYRVQNPKRLEAAYRETC 875

QY 881 SRRGTAAPVLLSGIYQVPSLSPDAMERNHRRGDELYLTPBAANWFEANKPAQVLYTT 940

Db 891 SRRGTAAPVLLSGIYQVPSLSPDAMERNHRRGDELYLTPBAANWFEANKPAQVLYTT 935

QY 941 EDTARTANALBELDATVEGRACAGCTISPGYHVOFTAGVSGSGSKSIQGGDVVVY 1000

Db 951 EDTARTANALBELDATVEGRACAGCTISPGYHVOFTAGVSGSGSKSIQGGDVVVY 995

QY 1001 PIRELNSWRKRGFAFTPTTAARVYIGRRVYIDEAPSLPHLLHMOARASSVHLGLDP 1060

Db 1011 PIRELNSWRKRGFAFTPTTAARVYIGRRVYIDEAPSLPHLLHMOARASSVHLGLDP 1055

QY 1061 NOIPALIDEHAULVPAIRBELAPTSWMXYTHRCPADVCELIRGAYPKIOTTSRVLKSLFW 1120

Db 1071 NOIPALIDEHAULVPAIRBELAPTSWMXYTHRCPADVCELIRGAYPKIOTTSRVLKSLFW 1115

QY 1121 NEPAIGOKLVYQAAKAAKAPGATVYHAGAGATFTTTATADARGLIOSSRAHAIYALT 1180

Db 1131 NEPAIGOKLVYQAAKAAKAPGATVYHAGAGATFTTTATADARGLIOSSRAHAIYALT 1175

QY 1181 RHTKCVILADAGLLREVGISDVIVNPFLLAGGVGXHRRSVYIRGNPNQNGTLOAPRP 1240

Db 1191 RHTKCVILADAGLLREVGISDVIVNPFLLAGGVGXHRRSVYIRGNPNQNGTLOAPRP 1235

DB 1176 RHTEKVEYIDAPGLRREVGISDAIVNNFLAGSIGHQREPSVIRPGNDPNDVTLAAPP 1235
 QY 1241 SCOTSAVHOLAEELGHRPAVAVLPPCPPELEGLLMPOLVSDSVLPELTDYHCR 1300
 DB 1236 SCOTSAVHOLAEELGHRPAVAVLPPCPPELEGLLMPOLVSDSVLPELTDYHCR 1295
 QY 1301 MAAPSORKAVLSTLVGRYGRRTKLYEAHSDVRESLARFPTIGPVQATTCETELVEAM 1360
 DB 1296 MAAPNOKAVLSTLVGRYGRRTKLYEAHSDVRESLARFPTIGPVQATTCETELVEAM 1355
 QY 1361 VEGGODSVALELDLCNRDVSRTTFKCKCNKFTTGTTIAGKVGQISAMSKTFCALFG 1420
 DB 1356 VEGGODSVALELDLCNRDVSRTTFKCKCNKFTTGTTIAGKVGQISAMSKTFCALFG 1415
 QY 1421 PMFRATKEITALLPPIFYGDAEESVFAAAGSACWVFENDFSEPTSTONNFTLGL 1480
 DB 1416 PMFRATKEITALLPPIFYGDAEESVFAAAGSACWVFENDFSEPTSTONNFTLGL 1475
 QY 1481 ECVYMEBCGMPQWLIRLYHLVRSAMILLQAPRESLKGFWKHSBPGTLLMNTVWMAAIIA 1540
 DB 1476 ECAIMEBCGMPQWLIRLYHLVRSAMILLQAPRESLKGFWKHSBPGTLLMNTVWMAAIIA 1535
 QY 1541 HCEFRFRRAAARKGDSVYLCSYRSRMAALLIACCGIKLYDVPYIGLYAGVVAAPG 1600
 DB 1536 HCEFRFRRAAARKGDSVYLCSYRSRMAALLIACCGIKLYDVPYIGLYAGVVAAPG 1595
 QY 1601 LGTLPDVVRFRAGRLSEKMGPGPERABQLRLVACDFLGLTNVAQVVDVVSRYGVSPG 1660
 DB 1596 LGTLPDVVRFRAGRLSEKMGPGPERABQLRLVACDFLGLTNVAQVVDVVSRYGVSPG 1655
 QY 1661 LVHNLIGMLQTIADGKAHFTETIKPVLDLTNSIIQRYE 1698
 DB 1656 LVHNLIGMLQTIADGKAHFTETIKPVLDLTNSIIQRYE 1693
 RESULT 5
 ID RPO_CGWS STANDARD: PRT: 1648 AA.
 AC P19523; Q83208; P89877; P90356;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
 DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)].
 OS Cucurbit green mottle mosaic virus (watermelon strain SH) (CGMYV), and
 OS Cucurbit green mottle mosaic virus (watermelon strain W) (CGMYV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 OX NCBI_TaxID=12236, 12237;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SH;
 RX MEDLINE=91311400; PubMed=1856687;
 RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
 RT Sato T., Motoyoshi F., Nishiguchi M.,
 RT "The complete nucleotide sequence of cucumber green mottle mosaic
 RT virus (SH strain) genomic RNA."
 RL J. Gen. Virol. 72:1487-1495(1991).
 RN [2]
 RP REVISIONS:
 RC STRAIN-SH;
 RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
 RL Sato T., Motoyoshi F., Nishiguchi M.,
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1496-1648 FROM N.A.
 RC STRAIN-W;
 RX MEDLINE=89073773; PubMed=3201760;
 RA Saito T., Imai Y., Meshi T., Okada Y.,
 RT "Interviral homologies of the 30K proteins of tobamoviruses."
 RT Virology 167:653-656(1988).
 CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA

CC CAPING AND AN RNA HELICASE.
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN
 CC CODONS FOR LYS-1144 AND GLN-1146.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch.)
 CC
 CC EMBL: D12505; BAA18895.1;
 CC DR EMBL: D12505; BAA18896.1;
 CC DR EMBL: J04322; AAA46382.1;
 CC DR PIR: J01157; WMTMS2.
 CC DR InterPro: IPR000606;
 CC DR InterPro: IPR001788;
 CC DR InterPro: IPR002588;
 CC DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
 CC DR Pfam: PF01443; Viral_helicase1; 1.
 CC DR Pfam: PF01660; Ymethyltransf; 1.
 CC KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 CC FT CHAIN 1 1648 RNA-DIRECTED RNA POLYMERASE.
 CC FT NP_BIND 863 870 METHYLTRANSFERASE/RNA HELICASE.
 CC FT NP_BIND 1 1144 ATP (POTENTIAL).
 CC SQ SEQUENCE 1648 AA; 186549 MW; 1D1AFEEB7B65595 CRC64;
 Query Match 4.5%; Score 406.5; DB 1; Length 1648;
 Best Local Similarity 18.8%; Pred. No. 6,7e-18;
 Matches 350; Conservative 220; Mismatches 633; Indels 661; Gaps 73;
 QY 77 CLEVGA-HPRSDNDPNVLCRC---FLRPVGRDQVQWYSAIPRGPAANCRSALGLP- 130
 DB 105 CYDIGNYTOHLEFKGSYVNCNCPCLDLKAVARNV-YNDMTTOHYORHSGSGCRPLP 163
 QY 131 ---PADRTY-----CFDGFSCAFMAETG---VALYSLHDLMPADVAMARHGX 174
 DB 164 FQIDAFRRYDSSPCAVTCSDFQECSDYFSGRDNHNAVLSHSYIDIPYSSIGALHRKNV 223
 QY 175 TRIYALHLRPEVLL--PRTYITTSYLLIHDDRAVYVTEGTSAGYINVDVILRAWIR 232
 DB 224 RVCYAAHFSEALLLSSPVGNLSIGAQFRVDDDVHFLFSEESTLHYHSLLENILYV 283
 QY 223 TRKIVDHPYIER---VRAIGCFVYLLTADEPSPMPVPPYPRSTEVYVRSIFPGGS 289
 DB 284 RTFPPADDRVYIKEMVAKVVDIFFRLVADTHMLHKSIGHYSKSKSEFAL-----NT 338
 QY 290 PSLFSPACSTKSTF-----HAPVHW 311
 DB 339 PRIF-----QDKATFSVWFPEAKRKVLIPKPELSRPLSGNVKISRMLVDADFVHTIINH- 393
 QY 312 DRMLGEGATIDDOAFCCSRMTLRLGISTYVYVYGVYALVANGMNASDALTAHTAULTI 371
 DB 394 -----STYDNKALVKNQVSEVRSIRSVIVGVYSEVMVVPVQDLDISFISFLV 446
 QY 372 CHQRYLRTQAISKMRRLGVEHQAOKFTIRLYSWLFEKSGNDYIPRGLOLYAACRMLSA 431
 DB 447 -----KVRKVOIE----- 454
 QY 432 GFHLDRVLVFEDESVPCRTFLKLVAGKFCPCFMRMLGDECTCELPBAGLVGDHGDNE 491
 DB 455 -----LMSDKVYIEAAGLLRRFADSLKSAVEGLG-DCV-----YDALVGTGWFDT 499
 QY 492 AYGESEYDAPEPAHLDVSGTYAVHGHOLEALYALVNPQDIARASRLATVVLVSPDR 551
 DB 500 SDELKVLPL-EP-----FMFESDYLEGMVEA-----DAKIRRESVS-BILASGDD 542
 QY 552 L-----ECRTVLGNKTER-----TVVQGHILDEANGPQOYVSPFASQSMG 593
 DB 543 LFKKIDIRNINYSGVFEVDEKFOEPCKELNVNPMILIGHYEA-----IFSQAGVTVTG 596


```

OY 1138 -----ANPGAI-TVHEAGATFETTTI-ATAADRGILLOSSRAHAIVALTIRHTEK- 1185
DB 1018 LKRGYEVEVSTFGEINIVHIOGETFEDVSVRLTPRLPELISSSPHVIALTRHTKSE 1077
OY 1186 ---CVILDAGGLREV-----GISDVIVNNFLAGSEVGHRSVIRPGNPDNLGLQA 1237
DB 1078 KYTSVVLDP---LVKWSDSLKVSDFLLDMKVADAGLLXDLQVSGTRKG---ENL----- 1126
OY 1238 FPSQOISAYHQAELGHRPAR-VAAVLPPCELEOGLLYMPELTVSDVLFELTDI 1296
DB 1127 FVP-CPRSGYISDMQYVYDFLVPGNSFILNEDYAVTMNL--RENNLVKVCOTIDFSKS-- 1181
OY 1297 VCRMAAPSORKAVLSTLVCRVGRKTKLYRAHSDVRESIARLPV--IGPVQATTCELY 1354
DB 1182 ---VSVPRQOOEFTFP-----AHRTAERPRRSAGLEENLVAMIKRNFSPDLTGILDI 1232
OY 1355 ELVEAWERKQDSAVLELDCNRDVSRIFEFOKXC---NKFTTGENI----- 1399
DB 1233 DTALVYNNKWDAYIIDLSGVNTPTSDAFHMMAKOEKSTIGOLADPDFVDLPAIDQ 1292
OY 1400 -----AHGRVGGGISA-----WSKTFCALFGFWFAIEKEILLALPPN--IF 1439
DB 1293 YKHMIKQPKOKIGLSPQDEYALQTIYHSHKOINAFGLFSELTRQLERIDSSKFLF 1352
OY 1440 YGAYEESV--FAAASGACSCWFFENDESEPDSTONNESLGLCEVVMESGMPWLIRL 1497
DB 1353 YTRKTPQIIEAFSDLDSTVPMVELDIDSKYKSONEFCAVEYLIMKLGINGFL-- 1409
OY 1498 YHLVRSAMILQAPRESLKG-----WKHSGEPTGLMTVMMAIILAHCEYERDF 1548
DB 1410 -----EEVWKQGHKRTSLKDYTAGIKTCLMYQKSGDVTTTIGNTVITIACLASIKIPMDKV 1465
OY 1549 RVAAFKGDSDVVLCS---DYRQSRMAALJAGGLKLKVDYR-IGLYAG-VVVAPELGT 1603
DB 1466 IKAFCGDDSDILVTPKGLDIPDIOSGANLWMEFAKL--YRRKYGVFCRYYIIHHRGA 1522
OY 1604 LP--DVVRFGRLESEKRWGSGPERAEQRLAVCDPLGLGNVAOV--CVVVSRYVYVSP 1659
DB 1523 IYVYDPLKLKLGCKKI-KSLDLEFRISLSDVSSSLNCAVFGOLDALAEVH---- 1577
OY 1660 GLVHNLGMLQTIADGKAHFTETIKPYLD 1688
DB 1578 -----KTAVNGSFAFCSTVKYLSLD 1596

RESULT 8
POLR TYMV STANDARD; PRT; 1844 AA.
AC F10358;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).
OS Turnip yellow mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
OX NCBI_Taxid=12154;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88289359; PubMed=3399308;
RA Morch M.D., Boyer J.C., Haenni A.L.;
RT "Overlapping open reading frames revealed by complete nucleotide
RT sequencing of turnip yellow mosaic virus genomic RNA.";
RL Nucleic Acids Res. 16:6157-6173(1988).
CC -I- MISCELLANEOUS: The 206 KDA PROTEIN IS POTENTIALLY A POLYPROTEIN
CC (BY HOMOLOGOY TO THE LONGER PROTEIN OF THE SINDRIS VIRUS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: X07441; CAA30322.1; ALT_SEQ.
DR PIR: S01956; S01956.
DR MEROPS: C21.001; -.
DR InterPro: IPR006066; -.
DR Pfam: PF01443; Viral_helicase1; 1.
KW Transferase; RNA-directed RNA polymerase; polyprotein; ATP-binding.
FT NP_BIND 976 983 ATP (BY SIMILARITY).
FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).
SQ SEQUENCE 1844 AA; 206640 MW; A016D758C83D128C CRC64;

Query Match 3.9%; Score 349; DB 1; Length 1844;
Best Local Similarity 20.6%; Pred. No. 3.6e-14;
Matches 385; Conservative 203; Mismatches 649; Indels 634; Gaps 93;

OY 157 LHD-----LMPADVAEMARH-GXTRLYAALHLPPEVLPRG-----TYHTSYLLIHD 204
DB 141 MHDALMYHPSQIMDLFLRKPNLERLYASLVPEPAHLSDQSFYPKLYVTYTRHLLHY- 199
OY 205 GDAVVYTEDTSAGYNHDSILRAMITRKI-VGDHPLIEFVRAIG-CHFVLLTAAP 262
DB 200 -----VPEGHEAGSYNP-SDAHSMIRINSILGNHHLVTLIESGVPYHSLIORGP 252
OY 263 EBSPMYVYPRSTEVYVNSIFGP----- 286
DB 253 PPDSIQAPPTLMTSDLFRSYQEPRLDVVSFRIPDAIELPQATFLOQPLRDLVPRAVYN 312
OY 287 -----GGSPLFPASCSKSTGFHAVVHIMDLMLFGATLDQACCSGLMT 333
DB 313 ALTTYRAVNTLFTSDPAAVVRHSSKPDHDWTSNAMDVLQTF-----ALLNV 361
OY 334 YLR-GISYKVTG-----ALVANEWNASEDALTAITAAYLTICHOYRLTOAISGMR 387
DB 362 PLRPNVYVHLOSPIASLSLYLQHRRL--TATAVPILSFLLL-QREL----- 408
OY 388 RLCVEHNAQ-KFTIRLYSWLEKSGRDYIPGROLQFAQCRMLSAGFHLDPRVYDESY 446
DB 409 PLPIPLAEVKSITAFRELYRKKE---PHNPLDV-----FHLOHRYVNHSAI 453
OY 447 PCGCRFFLKVKAGPCCFMMLQOECCLFEPABGLVGDHNDNEAVEGSEVPAPPAHL 506
DB 454 -----SAVRPASPPHQ 464
OY 507 DVSGTYAVHGHOLEALYRALNVPDIAARA-----SRLTATYELVASPOR--- 551
DB 465 KLP-----HALQKALLLLRPIPLTATPFFRSEOKSMPLNALSMTLKRFLPMQASL 519
OY 552 ---LECRIVLGNKTRFTYVVDGAHLEANGPEQYVLSFDSRQSGKAGSHSLTYELTPAG 607
DB 520 VLLALSESSILHLKRLSPPTLQAH-----DTYHRLHPSYSIQWERTP-- 564
OY 608 LQYRISNGL---DCTATPPGAPSAAGEVAAPFASALRYRYKRFQHSLSLTGMLHPE 664
DB 565 LSLPRTTALFPTPTTSTAPPDRSEASLP---PAFAST-----FVPR----- 603
OY 665 GLLGIFPP---FSPG---HWESANPECGGLTYTRTWSTSGF-----S 702
DB 604 -----PPRAASSPGAOPPTTTAAPTPIEPT--QRTHONSIALASSSTTEPPPPINS 655
OY 703 SDFSP-----PEAAAPMAATPGL-PHSTPPVSDIWLPPPEE----- 740
DB 656 PDMTSPAPVLPFINSPPRRPPQLPATRPDLERAHNPPLISIPHQDPTDSADPLMGSHLLH 715
OY 741 FOVDAAVPVPAAPALPGRVYLTPPPPV-----HKPS-- 775
DB 716 HSLPAPPTHLPSSQLLPAPLTNDPPAIGVLPFEELHBRYPENTATETLTRLSPSHNH 775
OY 776 IPPSPNRRLLYTPPGAVAYAGSLPESQDMVLVN--SNPGHRCG-----GLCH-- 824
DB 776 LPOPTLN-CLLSAVSQOTYVSEHLMESIQTLIPDSQLSNEENTNIGLSTENHTALAHLY 834
OY 825 -----AFYORFPEAFVPTFEIMREGIAAYTLPRPIIHAVAPDYRVEONPKRLEAVRET 879

```

```

Db 835 NFOATVYSDRGPIILFSPDITK-----IDTHTTGPSHSPG-KRLGSP--881
Qy 880 CSRRGTAAYPLLG-----SGIYQVPSLSPDAMERNHNPDELY---LTPAAMFEA 929
Db 882 -SAKHGSPDLIRAMKSPFKSVMY-LPFS-----EAHNPTSISAKNLISMKNQFPG 933
Qy 930 NKAQAVLTITETARTANALEIDATEYGRACAGCTISPG-----IVHQFAGVPGSG 965
Db 934 VLSLDVSTGQRTGPPKERTIQIDHYLDT-----NPGKTPVVFH---AGEPAGCG 981
Qy 986 KSRISIQ-----GDVVVYVPTRELNSMR---RGPAA---FTPHTAARYTIGRGV 1032
Db 982 KTYPIQGLKTKFKDPR-VSCPTTELKTEWKAMELHGSQSRFTWESSILKSRILV 1040
Qy 1033 IDEAPSLPRLHLLHMOASSVHL---LGGP-----NOIPAIDEHGLVP 1075
Db 1041 IDEIYKMPKGYLDLSILADPALELVIILGDPDQGEYHSGSKSDSNHRLPS---ETLRLLP 1097
Qy 1076 AIRPELAPTSMMXVTHRCRADVCEL-----RGATPRIOI-----TSRVLRSLF 1119
Db 1098 YI---DMTCWM---SYRIQCIARLFQIHSEFNAMQVIGSVSTPHQSPVLTHSHASSLT 1151
Qy 1120 WNEPAIGOKLVYTOAKAANPAGATVHEAOGATFETITATADARGLIOSSRAHAIVAL 1179
Db 1152 FN--SLGVR-----SCTISSQGLFCOPPAITVLDNYTKWLS--ANGLVAL 1194
Qy 1180 TR-----HTKCVILDA--PGLREVGISDIVVNNFL 1210
Db 1195 TRRSRGVQEMGPSSVYGTINGSSAMFSDAFNNSLIIMDRFSLPQL---KLITSPLTT 1251
Qy 1211 AGGEVGHKRP-----VIRGNPDQMLGLQAPPS 1241
Db 1252 RGKLGATPSPASPTHRSPFHLPHIPLSYDRDEVTVNPTLDQGEYERLDT--HFLRP 1309
Qy 1242 COISAHQIAELGHRPAVAALPPCELEOGLILYMPOLVTSDSVLYEL-----T 1294
Db 1310 SRLPLHFDLPAL--TPPYVTSVDP-POAKASPVY-PCGF--FDLSIAFELPAHDPSTR 1363
Qy 1295 DIVHCKMAAPS---ORKAVLSTLVGRYGRKLYEAAH-----SDVRESLARPT 1340
Db 1364 EIIHKQSSQOFPWFRPFSLSL---QPSLLISAKHAPNHDPPLLIPASINKRL-RFR 1416
Qy 1341 PTIGPVQATP-----C-----ELYELVEMVKGQDGSAYLE 1372
Db 1417 PDSPPHOITADVVYLGLOPHSLICRAYSROPNSTVPPNPDELFAECLSLNEYAOLSKTOS 1476
Qy 1373 LIDLNRDVS-----RITFEOKCKNFTTGETIANGKVGQGISAMSKTECALFGP--W 1422
Db 1477 TIYANASRSDPDRHTTIVKIFAKAOKKVNDSIFGSMKACQITLALMHDYIYLVGPYKKY 1536
Qy 1423 FRAIEKELLALPNNIF-----YGDAYEESVFAAASGAGSCNVEFNDESEDDSTQ 1473
Db 1537 ORIFEDN---ADRPNIYSHGKTPNOLRDMCOEHL-----THSPKIANDYTAEDSQ 1586
Qy 1474 NNSLGLCECVMEGCMQMLRLY-HIVRSANILQAPKESLGEFKKHSGEGCTLLMNT 1532
Db 1587 HGESVLEAKMKRLNPISHLQDLHVLKTNVSTQFGLTCM-----RLTGEGETYDDNT 1641
Qy 1533 VNMMAIHAHYERFDRVAFAKGDSDVVLCSDY--ROSRAAALIGAGCGLKLVDY---- 1586
Db 1642 DYVLAVIYSGYDVGSCPIW--SGDDSLI---DHPLETRHWPVSLKRLHRLFELELSHP 1697
Qy 1587 -----RPIGLYAGVVAAGLGLTLPD---VVRF--AGRISEKWNQGPGR 1625
Db 1698 LFCGYVVGAGCIRNPLALFCKIMIAVDDALDRLRLSYLFEFTTGHLGESSIMHLPER 1757
Qy 1626 ABOLRLAVCDF 1636
Db 1758 HVOYQASCEDF 1768

```

RESULT 9

```

POLR_TYMC
ID POLR_TYMC STANDARD; PRT; 1844 AA.
AC P28477;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).
OS Turnip yellow mosaic virus (isolate TYMC).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
OX NCBI_TaxID=31751;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92119261; PubMed=1731998;
RA Dreher T.W., Branson K.L.;
RT "Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a
RT cDNA-based clone with verified infectivity."
RL Plant Mol. Biol. 18:403-406(1992).
CC -1- MISCELLANEOUS: THE 206 KDA PROTEIN IS POTENTIALLY A POLYPROTEIN
CC (BY HOMOLOG TO THE LONGER PROTEIN OF THE SINDBS VIRUS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X16378; CAA34415.1; -
CC DR PIR; S19152; S19152.
CC DR InterPro: IPR000606; -
CC DR Pfam: PF01443; Viral_helicase1.1.
CC KW Transferase; RNA-directed RNA polymerase; Polypeptide; ATP-binding.
CC FT NE_BIND 976 983 ATP (BY SIMILARITY).
CC FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).
CC SQ SEQUENCE 1844 AA; 206612 MW; 02CB928FCCCA5EAL CRC64;

```

Query Match 3.8%; Score 346; DB 1; Length 1844;
Best Local Similarity 20.5%; Pred. No. 5,6e-14;
Matches 380; Conservative 210; Mismatches 660; Indels 608; Gaps 94;

```

Qy 157 LHD-----LMPADVAEAMARH-GXTRLYAALHLPREVLPPG-----TYHTSYLLIHD 204
Db 141 MHDALMAYHPSQIMDLFLKRPNELERLASLVPEEHLSSQSTPKLYTTTRHTLYL- 199
Qy 205 GDAVVYTEGDTSGAGYNDVSIIRAWITRTKI-VGDHPIYIERVRAIG-CHFYLLTAAP 262
Db 200 -----VPEGHAGSYNQP--SDAHSWLRINSIRLGNHLSVTLLSESGPYHSLIQGTP 252
Qy 263 ERSPPMRYVP-----YRST-----EYVNSIFG 285
Db 253 PPDPISLQAPPTLMAASDLFRSYOEPRLDVVSFRIPDALELQOAFLEOPLRDRVLPRAYVN 312
Qy 286 P-----GSPSLPSPASCTSTFHAVPVNHIMDMLFGATLDQOACCSRLMT 333
Db 313 ALFTYTRAVTLRTSDPAALVVRHSSKRPDHDWTSNAMDNLQTF-----ALNV 361
Qy 334 YLR-GISYKYTVG-----ALVANEQWNASSEDALTAAXTAAYVLTIQHORYLTOAISGMR 387
Db 362 PLRPNVYVHYLQSPIASLSLYRQHMRL--TATAVYILSFLLL-QRFL----- 408
Qy 388 RLCVEHNAQ-KETIRLYSWLFEKSGRDYIPGRQLOFYAQCRRMLSAGFHNDPRVLVEDEV 446
Db 409 PLPIPLAEVKSITAFRELRKKE---RHHPRDV-----FHLOHRIIRNYSIAI 453
Qy 447 PCRCRFFLKVAARFCFCFMWMLGOECTFLERAGLVGDHGHNEAESEGVPAEPAHL 506
Db 454 -----SAVRPASPPQ 464
Qy 507 DVSQTYAVHGHQLEALYRALNVPQDIARA-----SRLTAVVELVASPDR--- 551
Db 465 KLP-----HALQKAALLLRLPISPLTATPFPFSEQSKMLPNDELSTWTKRFPALPQASL 519

```

```

QY 552 ----LECTVGNKTFRTTVVGDHLEANGBPQVYLSFDSARQSGASHLTVELPAG 607
Db 520 VLLALSESSILHKLFPPTLOAQH-----DTHRHHLHPGSSYLQWERTP-- 564
QY 608 LOVRSSNGL---DOTAFPPPGGASAPAGVAAFCALYKNTTOR---HSLTGLW 660
Db 565 LSIPTTVAFLPPTTSTAPDRSEASLP---PAFAST-----FVPPPPAASSPGA-- 613
QY 661 LHPGLGIFP-PFSPGHIMESANFCCEGLTYRTWSTGSSDFSP-----P 708
Db 614 -QPTTTAAPPPTPIEPTRAIQNSDLAESSTSTPPPPPIRSPTMTSAVYLPETINSP 672
QY 709 EAAPAMATPGL--PHSTPPVSDIWLPPSEEFQVD-----AAVPPAPD 753
Db 673 RRPPTLPATDLEBAHPPPLS-----IPHODPTSDVPLGSHLHSLAPTHPLPS 728
QY 754 PAGLPQVVLTPPPPPV-----HKPS--LPPSRNRRLTYT 788
Db 729 SOLPAPLTPDPTALGPPLPEELHPRYPENTATFLRLSLPSNHLPOPTLN-CLLSA 787
QY 789 YPDGKAVYAGSLFESDDCWLVA--SNPGHRPG-----GLCH-----AFQREPE 832
Db 788 VSDOTKXSEELHMSLQTLIPDSQLSNEETWTLGLSTHLTALAHLYFOATVYSDRGPI 847
QY 833 AFYPTFEIMREGLAAYTLTPRIIHAVADYVEQNPRLBAAYRETSGRGTAAVPLG 892
Db 848 LFGSDTIKR-----IDITFTGPPSHFSPG-KRLGSGP---SAKGHPSPDLIR 893
QY 893 -----SGIYQVYVSLSFDAEMRNHRGDELY---LTERPAANFEAKRPQVLTITED 942
Db 894 AMKSFYVSGN-LPPS-----EAMNFTSISHAKNLSNNKNGDGLSLDVGSTGORT 946
QY 943 TARTANLALTEIDAATEVGRACAGCTISPG---IYHOFTAGVPGSGRSIQO----- 992
Db 947 GPTKFERIQIDHLYDT-----NGKTPPVHF--AGFAGCGCTTYIQQLTKKL 994
QY 993 -GDVAVVVPRLRLNSRR---RGFAA--FTRPTAARTVIGRRVVIDEAPLPHLL 1045
Db 995 FKDR-VCSPTELTETKMTAMELHGSQSMRFNWESSILKSSRLVDELYKKMRGLD 1053
QY 1046 LHMGRASSVHL---LGDV-----NOIPADFEHAGIYPAIRELAPTSWMX 1088
Db 1054 LSIADPRLVLIIDGLDGEYHSQKSSNHRPLS---ETLRLPIT---DMYCW- 1105
QY 1089 VTRHCPADVCELI-----RGAYPKIQT-----TSVRLSLFENEPAGCKLXYT 1132
Db 1106 -SYRLPQIARLFOIHSFNAMQVIGSVTFHDOSPVLTNHASSLTFN--SLGVR----- 1158
QY 1133 QAKAANAGAITVHAQCATETETITATADARGLIOSSRAHAIVALT- 1181
Db 1159-----SCTISSQGLTFCDPAIIVLDNYTKWLS--ANGVALTRSRSGVQFMGPS 1207
QY 1182 -----HTEKVIIDA--PGLIREGISDVIVNFFLAGGVGXHRP--- 1220
Db 1208 SYVCGTSSAMFSDAFNNSLIMDRYFSLPOL---KLITSPLTTRGRLNGATPSAS 1264
QY 1221 -----SVYPRGNPDONIGTLQAPPPSCQISAYHQLAEEL 1254
Db 1265 PTHRSPPNHLPHPLIPLSYDRDFVTYNSTLIPDQGPETRLDT--HFLPERSRLHLHDLPAI 1322
QY 1255 GHRAPAAVLPPELEQGLLYMPQELTVSDVLYFEL-----IDIVICRAAAS- 1305
Db 1323 -TPPVSTVSDP-POAKASPVY-PGEF--PDSLAAPFLPAHDPSTREILHKDDSSNOFP 1376
QY 1306 -ORKAVLSTLVGRYGRRTKLYEAAH-----SDVRSRLARFTPTIGPVQATT--- 1350
Db 1377 WEDRPFSLSLSC-----QPSLSIAKHAHPNDPPLLPAISINRL-RFRSDSPHOITADV 1429
QY 1351 -----C-----ELVELYEAAMEVKGQDSAVLELIDLCNRDVS----- 1381
Db 1430 VLGQLFHLSLCRAVSRQPNSTVPPNPELFAFCISLNEYAQJLSSTQSTIVANASRSDPW 1489

```

```

QY 1382 ---RTTFQKXCNKFTTGETTAHGKVGQGISAMSKTFCALFQ---WFAIEKEIALLP 1435
Db 1490 RHTTVKIFAKAHKYNQDSIFGSKACQTLAMDYVILVGPYKKYORIPDN---ADRP 1546
QY 1436 PNIF-----YGAVEESVPAAYAGSAGSCWVFENDESEPSTONNFSIGECVME 1486
Db 1547 PNITSHCGKTNOJLRDWCQEHNL-----THSTPKIANDYAFQSOHGESVLEALKMK 1599
QY 1487 ECGMPOWLIRY-HLYRSAMTLOAPKESLKGFMKXHGSEPTLLMNTVMNAIIAHCEYF 1545
Db 1600 RLNIPIHILQIAHVKHTVSTQFGLTCM-----RLNGEPTDYDNDNDYNAIVISQDV 1654
QY 1546 RDRVAAFKGDSVYVLSQV---KOSRNAALIIACGKLKDY----- 1586
Db 1655 GSCPIYM--SGDSDLI---DHLPLTRHDMPVYLKRLHFKLELASHPLFCGYVGPACGI 1710
QY 1587 -RPIGLVAGVYVAVGLGLTP-----VYRF--AGRLSKNMGPPPEREQLRLAVCF 1636
Db 1711 RNPLALFCKLMAIVDDALDRLSLYTEFTTGHILGSLMHLLEPHVYOVSACFDF 1768

```

RESULT 10

RRPO_ORSYVC STANDARD: PRT: 1612 AA.

```

AC P9659: 039640:
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Odontoglossum ringspot virus (Isolate Korean Cy) (ORSV-Cy).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138661;
RN [1]

```

RP SEQUENCE FROM N.A.

```

RX MEDLINE=96381046; PubMed=8789059;
RA Ikegami M., Isumura Y., Matsumoto Y., Chataani M., Inoue N.;
RT "The complete nucleotide sequence of odontoglossum ringspot virus (Cy-
RT 1 strain) genomic RNA."
RL Microbiol. Immunol. 39:995-1001(1995).
CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR LEU-1112 AND GLN-1114.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D13941; BAA21828.1;
CC DR EMBL: S83257; AAB49498.1;
CC DR InterPro: IPR000606;
CC DR InterPro: IPR001788;
CC DR InterPro: IPR002588;
CC DR Pfam: PF00978; RNA_dep_RNApol2; 1.
CC DR Pfam: PF01443; Viral_helicase; 1.
CC DR Pfam: PF01660; Vmethyltransferase; 1.
CC KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
CC FT CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.
CC FT CHAIN 1 1112 METHYLTRANSFERASE/RNA HELICASE.
CC FT NP_BIND 826 833 ATP (POTENTIAL).
CC SQ SEQUENCE 1612 AA; 183252 MW; 3F803A2B9611E0DF CRC64;

```

Query Match 3.88; Score 339.5; DB 1; Length 1612;
 Best Local Similarity 19.28; Pred. No. 1.2e-13;
 Matches 351; Conservative 232; Mismatches 659; Indels 585; Gaps 85;

```

OY 94 LHRCE-----LRPYGRDQWYSAPTRGPANCRSRLRGLP-----PADRTYC 137
DB 123 VHCPEMLNDIRVAHNIQDQTVSY--LAKLRSK--KGLPVFQOGAFKNMNDPVAVCC 179
OY 138 PDGFSRCAAEET-----GVALYSLHDLMRPAD--VAEMARHGCTRLAALHLPPEVLPP 191
DB 180 DKRFQOCSSVDLPCKTAVVALHSTYDI--PADEFGAALLRKVHCYACAFHPSENLL--- 236
OY 192 GYVHTTSTYL-----IHGDRAVVTYEGDTSAGYNHDVSLIRAMIRTT-----KIYG 238
DB 237 ---ETTSAPLDEIGATFYKSGDRLSFEFFONESTLINEHSYKVKYCKTFEPASNRFY 293
OY 239 DHPRIEYRAIGCHPVLLT-----AAPESPMPYV----- 271
DB 294 HKEFMCTRVNWFCKTKVDYTFLEGRVYTRGDSQOFTANDEAMEYKTLAMKCEPT 353
OY 272 -----YPRSTEVVRSIF--GPGGSPSLFSPACSTKTF--HAVVHIWDRMLFG 318
DB 354 IFRDRAAVWFMPKYKDMVIVPLFDGSVTSCKMKRSEVWVNDVYTVLNIH----- 405
OY 319 AFLDOAFCCSLMTYLRGISTYKTVGALVANEGMNASDALTAIXTAAYLTICHOYLR 378
DB 406 RYQOKKALYKKNVLSFVESIRSRVILINGVASRSEMDVDSKYLQA--LSMTFL-----LO 457
OY 379 TQAIKGMRLGVEHAKQF---ITRLXSWLFKESGRDI--PGROLQFYAOCRRMLSGFHL 435
DB 458 TT--LBAKQVYVKKFKOKFDVTNLEWKOISDAVGDLPESTK-----ETLISGGF-- 507
OY 436 DRYVLVDESVPCRCRTFLKAYAGKFCCEPMRLGOECTCFLEBAELVGDHGHNEAYEG 495
DB 508 ---VVAEGSLDI-----KTRPEYITF-----ADKLWME-----YKA 536
OY 496 SEVDPAEPAHLDVSGTYAVHGHOLEALYNPODIAARASRLIATVELVASPDLECR 555
DB 537 TE---ELQHLDISKPL-----ERAEKYNAL-----SELSTYKLE--COBFDITQFK 577
OY 556 TYLGNTKTFRTYV-----DGAHLEANGPEQVYVLSFDASQSGAGASHLTVELTPA 606
DB 578 NICEEKDIDPPVAKYIYAIMKNELTLPKNPPTPEALS--DA-----LSPLEK 623
OY 607 GLQVRISNGLDCTATFP-----PGAPSAAPGEVAAFCALRYRNRTQSRHSLTGJLM 660
DB 624 DLDMRDLKLSTCAFPFVKTLDSGLPKQSYGDERQF-----ESQSVVSVSD 672
OY 661 LRPBGLIGTFPPSPCHIMESANPFCGEGTLTTRTWSSTGFSFSDPPEAARAPMATGP 720
DB 673 FHLKSY-----ESYKMKSMSSAVYTGPLKVOOMKNYM--DYLSSASISAT-- 714
OY 721 LPHSTPVPADIVLPPSEEFQVDAPVPAPDPAGLPGPVVLTTPPPPVHKSIPPS 780
DB 715 VSNLCKVLADVYGADPESAE----- 734
OY 781 RNRRLTYTPDGAKYVAGSLFESDCDMLVNASNPGHPPGGGLCHAFYORFEAFYPTERT 840
DB 735 --KSGYDVYVK-----KWLKPK-----GKCHAN----- 757
OY 841 MREGLAAYLTLPPIITHAVALPDYEQNKRLBAAYRETCSSRGTAAYPLLSGITYQVPY 900
DB 758 ---GVAELNNGEKVIV-----LLEMA-----DGFICGD----- 783
OY 901 SLSPDMERNHRRGDELTYLE--PAANMFENKPAQPVLTITFEDTARTNIALEIDAARE 958
DB 784 -----WRVAVAVSSDSLISDGKLOTJLISLCKDGEPV----- 815
OY 959 VGRACAGCTISPGIYHOFYTAGVPGSGRSIQOG---DVDVYVVPVTRRLNSWRRG-- 1013
DB 816 -----PSDAKVTLVGDVPGGKTKELIETYNPDEDLIVGKRAKMIIRAK 864
OY 1014 -----FAFTPTHTAARTTIGRRVVIDEAPSLPPLLILH-----MORA 1051
DB 865 SGHVRATKDNVFTVSFLMHLKPKTY--NKLFIDEG-----LMLHTGCANFLIALSHC 915

```

```

OY 1052 SSVHLLGDPNOIPAD-----FEHAGLVPAIRBELAPTSMKXVTHRRPAD 1096
DB 916 REAMVEGDDEQLEPFINRVANFPYPKHFAVLVDHR---EVR-----RLSLCPAD 962
OY 1097 VCELLIGAAP--KIOTSRVRLSL-----FWN--EPALGOKLVXTOAKA----- 1137
DB 963 VYHFMNSKIDGKVLCTNDVIRSDAENVGKGKGFNFKSNPKLGIITTFQSDAKELKNG 1022
OY 1138 ---ANPGAI--TVHEAGATFETTTII--ATADARGLIOSSRAHAIYALTRHTEK---CV 1187
DB 1023 YEEVSFGEIINTVEHIOGEFFEDVSIVRLTPPLELISKSSPVALVALTRHTEKSFYYS 1082
OY 1188 IUDAPGLREVG---ISDVYNNFELAGGEVXHHPSVIRGNPNQNLGTQAFPPSQ 1243
DB 1083 VLDP--LVKVCSDLSKVSVDPLDMIVVDAGILQOLQVGSIFKG---ENL-----FVP--CP 1131
OY 1244 ISAYHOLAEELGHPAPVAVALPPELPEL---EQGLTYM---POLTVSDSLVLELDIYH 1298
DB 1132 KSGY-----ISDMQFYUPTLLPGNSTLINEYDAVYMNLRNNLNVKDCIDFSKS---- 1181
OY 1299 CRMAAPSOAKAVLSTLVGRGRTKLYEAHSDVRESLARFIPT--IGPVQATTCELYEL 1356
DB 1182 --VSVPROOQEEFTPYI-----RTAERPRSRGILENLVAMIKRNPSPDLTGILDEDT 1234
OY 1357 VEMAVEKGODGSAVLELDLCNDRVSRTTFQKXC---NKFTGETIAHGKVG----- 1405
DB 1235 AELVNVKFWDAYITIDELSGGNVTPTSDAFHRMAKOEKSTIGQLADFDFVLDPAIDQYK 1294
OY 1406 -----OGISAWSKTEFCALGPMFRAIEKILALLPPN--IFYG 1441
DB 1295 HMIKAQPKQKLDLSDPQDEYAALQTIYHKSQINAIIGPLSELTRDLLEIDSKFLFT 1354
OY 1442 DAYEESV--FAAASGAGSCWPFENDSEFDSQNNPSGLGECVWHEGCMQWILRLYH 1499
DB 1355 RKTPEQIEEFSDLDSTVMEVLELDISKYDKQNEFHCAVEYLIMEKILNGFL----- 1409
OY 1500 LVRSAMILQAPKESLKG-----WKKHSGEPGLMTNVTWMAAILAHCEFEFDPRY 1550
DB 1410 ---EYVKQGHRTSLNDYAGITACITCMYORKSGDVTTFIGNYITIAACIASMTIPMDKYK 1467
OY 1551 AAFKGDSDVYLS--DYROSNAALIAOGGLKLVDRP--IGLYAG--VVAAPGLGTLPP 1605
DB 1468 AAFCGDSDMLYIKGLDLPDIOGANLMMNFEAKL--YKRYRGYCGRIIHHNDGAIY 1524
OY 1606 --DVYAFAGRLSKNNGPPEBAEQRLAVCDPLRGITLNAOY--CYDVYSRKYGVSPGL 1661
DB 1525 YYDPVKLISLGGCKH--KSLDHLIEFRISLCDVYASASLNCAYYGOLDALAEVH----- 1577
OY 1662 VHNLIQMLOTIADGKAHFTETIKPVLD 1688
DB 1578 -----KTAVNGSFAFCSYKYLSD 1596

RESULT 11
RRPO.TOML STANDARD: PRT: 1616 AA.
AC P03587; 041352;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) (CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)).
OS Tomato mosaic virus (strain L) (TOMV) (TMV strain tomato).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
ON NCBI_TaxID=12252;
RX SEQUENCE FROM N.A.
RA MEDLINE=85157522; PubMed=6549393;
RA Ohno T., Aoyagi M., Yamashita Y., Saito H., Ikawa S., Meshi T.,
RA Okada Y.;
RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RT genome and comparison with the common strain genome.";
RL J. Biochem. 96:1915-1923(1984).

```

CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X02144: CAA26085.1; -
DR EMBL: X02144: CAA26082.1; -
DR PIR: A04195: WMTM8T.
DR InterPro: IPR000606; -
DR InterPro: IPR001788; -
DR InterPro: IPR002588; -
DR Pfam: PF00978: RNA_dep_RNAPol2; 1.
DR Pfam: PF01443: Viral_helicase1; 1.
DR Pfam: PF01660: Vmethyltransf; 1.
DR Transferrase: RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT NE_BIND 833 840 ATP (POTENTIAL).
FT SQQUENCE 1616 AA; 183564 MW; A8BC92985CFCAF CXC64;

Query Match 3.68; Score 322.5; DB 1; Length 1616;
Best Local Similarity 19.24; Pred. No. 1.5e-12;
Matches 349; Conservative 205; Mismatches 660; Indels 601; Gaps 79;

QY 91 PNYLHRCFLPVGR--DYGRKYSAPTRG--PAANCRSAL-RGLPRADRTYCFPGFSRCA 145
DB 128 PNLVDRIMHREGOKDIEYLSRLERGNKHVPFOKEAEFRVAVEMPREVYCHDTFOTCR 187
QY 146 FAAE--TG-----VALYSLHDLMPAD-VAEMARHGXTRLVYALHLPEVLLPFG--TYHT 196
DB 188 HSQECYGRVYALALHSTYD-PADFCGALLRKNVCHVCAAFHFSNLLLEDHVNIDE 246
QY 197 TSYLLHGDRAVYVYGDTSAGYNHDVSLIRANIRTKTVIGDHPVIERVRAIGCHFVL 256
DB 247 INACFOQDGRLETFSEFSESTLANSYSNLIKVCYCKT----- 284
QY 257 LITAAPSPMPYVPRSTREYVYRSIFGPGSPSLFPPSACSTSTFHAVPHIMDKML 316
DB 285 -----YFP-ASNRVYMKK-----FLVTRVNTW----- 306
QY 317 FGATLDDQAF-C SRLMTYL--RGISYKVTYGVALVANGWNASDALTAITAAVLTICH 373
DB 307 -----FCKSRIDTFLPKYGVANK-----GVDSEQFYKAMDANHYKKT--LAMCN 350
QY 374 -QRLKTOAISKG-----MRRLGYEHAOKFTRLYSWLFKSGR-----DYFGRO 418
DB 351 SEIRILEDSSSVNVWPFKMDM-----VIVPLFDSLSTSKTRKEVLVSKDFV----- 399
QY 419 LQFYAOCR--RWLSAGPHLDPRVLVPEESVPCRTFLKKVAGFCGFMRLGECFCFL 476
DB 400 -----YTVLNHRTTQAKALLTSNVLSFEVST--RSRVLIINGVTAR----- 438
QY 477 EPAEGLVGDHNDVAGESEVDPAPPAHLDVSGTVAVHGHQLEALYRALNVPODIARA 536
DB 439 -----SEMDVDKSLQSLSMTEFLH----- 458
QY 537 SRLTAIVELVAPDRLECRVVLGNKTRTIVVDGAH----- 572
DB 459 -----TKLAVLKDDLLISKRALGPKVYSOHWMEISLAFGAPPSIKERLINRKLKITE 513
QY 573 --LEANGPEQYVLSFD--ASROSGAGASHL-----TYELTPA--GLQVRISSGGLD 618
DB 573 --LEANGPEQYVLSFD--ASROSGAGASHL-----TYELTPA--GLQVRISSGGLD 618

DB 514 NALRIYVDLYVTFPHDLVSEYKMSVDMPLYDIRKKMEETEEMYNALSELVLKNSDKFD 573
QY 619 CTATPPPGAGAPSAAPGVAAFCSALYRNRTQGHSLTGGMLHPEGLGIFPPFSPCHI 678
DB 574 VDV-----FSQMCOSL----- 584
QY 679 MESANPCGECTLYTRTWTSGFSSDSFPPEAAPAMAAFGPLPHSTPPVSD--IWVLP 736
DB 585 --EVDPMIAKVIYAVVNSGSLTTFEQPLEANVALAL-----QDEKMSDGLAVYTSR 637
QY 737 PSEFQVDAAPVPAAPPAAGLPGPVYLTTPPPVYHKKSPSPSRNRLLYTPDCAKY 796
DB 638 DVEPDSIKSGMARGELQLAGISGVY-----PES-----SYRSEIE 674
QY 797 AGSLFEESDCMLVNASNPGHPPGGGLCHAFYQREPAEYPTPEIMRGLAAVYLTTPPII 856
DB 675 SLEQFH-----MATASSLIHQ--KCSIYTGGLKQOQKNFI--DSLVAASAAVSNL 724
QY 857 HAVAPDYRVEDQNPKRLAAVRETCSRGTAAVPLLGSIYQVPVLSFDAM--ERNHRPG 914
DB 725 VKILKDTAAIDLETROKFGVLDVASKR-----WLVPKSAKNHAGVETHARK 772
QY 915 DELYTE-----PAANPEANKPAQPVLTITEDTARTANLAIEDAATEVAGACAGT 967
DB 773 YHVALLEHDEFGIITCDNMRRAVAVSSSVYSDMAKLRTLRLLK-DGEPHVSSA----- 826
QY 968 ISPGIVHOFYAGVPGSGKRSIQOG--DVDVYVVPVTRELNSMRRR--GFAAFTPH 1020
DB 827 -----KVVLVDQVPGGCKTKELSRVNFEDLILVGRQAAEMIRRRANASGIIVATKD 880
QY 1021 TAARV-----TIG-----RRVVIDEAPSLPHLLH-----MORASSVHLGD 1059
DB 881 NVRTVDSFLMNYGARGCQRFLEIDEG-----LMHFGCVNFLVEMSLCIDAIVYGD 933
QY 1060 PNOIPALIDFHEAGLVPA--IRPELAPTSWMXVYHRCPADVCILRGAYP--KIOTSVLR 1116
DB 934 TQOIPYINRYVTFPPYHFAKFLDEVEYETRRRTLRCPADVTHFNQRIEGHVMCTSEKK 993
QY 1117 SLFW-----NEPAIGOKLYVXTOAKAA-----NPGATITVEAGCAFTETII 1159
DB 994 SVSGEMVGAASINPVSKPLKGLITFTQSDKEALLSRGADVATVHVEQGETYADVSLV 1093
QY 1160 -ATADAGLIOSSRAHAIVALTRHTK-----CVILDA-----P 1192
DB 1054 RLTPPTVSIILAROSPVALVLSRHTKSLKYTVMDPLVSIIRDLERVSSYLDMYKVA 1113
QY 1193 GLIREVGISDIYVN--NFLAGGEVGHXRPSTIYPRGNPDQNLGTOAAPPSCQISAHYQLA 1251
DB 1114 GTOXQLOVDVSFKFNFLVPAAPKTG-----DISDQFYDKC----- 1150
QY 1252 EELGHRPAPVAAVLPCEPELEOGLVYMPOLIVSDS--VLVFEELTDI-----VHC 1299
DB 1151 -----LPGNSTLLNNVDATYMKLTDISLWVKDILDMSK 1184
QY 1300 RMAAPSGKRAVLSTLVGRYGRFKLYEAASDVRESLARIPY--ICPVQATTCELYELV 1357
DB 1185 SVAAPKDKVKKPTLLPMV-----RTAAEMPROTGLEENLVAMIKRNFNSPELSGVVDIENTA 1239
QY 1358 EAMVEGODSAYVELDLNCRNDVSRITFPQKXCKF--TTGETIIGHKVG----- 1405
DB 1240 SLVYDKDFDSYLKEKKPKPKNFS--LFSRESLNRMIAKQEOVYITGLADPFDVLPAYD 1297
QY 1406 -----OGISAMSKTFCALFGPWRARKEITALL----- 1434
DB 1298 QYRMIKAOPKOKULDISIQTEYPALQITIVYHKKINAFIPLFSELRLOLIDSISRFL 1357
QY 1435 -----PPNI--FYGDAYESVRAAVSAGSGSCWYFENDSESDSTONNLSLGEVYME 1486
DB 1358 FTRKTPPAQIIEDEFGD-----LDSHVPMQVIELDVSKYDKSNEFHCAVEYEIWR 1407
QY 1487 ECGPQWLIRLHYVRSAMWLOAPKESLKGK-----MKHSGSEGPLMMNTVNMMA 1537
DB 1408 RLGEDPFLAEV-----WKQGHKRTYLLKDYITAGIKTCLMAYORKSGDVTTFTIGNTVITAS 1460

QY 1538 IIAHCEPDEFERNAKGDSDVYL-----GSDYRQSRNAALAGCGLK-LKYDRIPIGLY 1592
DB 1461 CLASMLPMELKLGAFCGDSDLLYPEKGC-EXPDIQQAALMMNMFAPKLEKKQY---GYE 1516
QY 1593 AGVVA---PGLQTLDPVDFVAFGRLE---KNMGPCPEAEOLRLAVCFELRGLTVAOV 1646
DB 1517 CGRYVHHHRCGLVYDPLKLSKLGAKHKKDM---DHLEFRRLCDVAESLNACAY 1572
QY 1647 CV-DVYSRYGVSP 1659
DB 1573 TQLDVAQGEVHTAP 1587
RESULT 12
RRPO_TOMSL STANDARD: PRT: 1616 AA.
ID Q9YKD6: Q9YJ37;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tomato mosaic virus (strain S-1) (TOMV).
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138314;
RN (1)
RP SEQUENCE FROM N.A.
RA Zhou X., Xue C., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence of a Chinese isolate of tomato mosaic
virus".
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
CC - FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
RNA REPLICATION.
CC - FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CAPING AND AN RNA HELICASE.
CC - MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL: AJ132845; CAB36997.1; -;
DR EMBL: AJ132845; CAB36998.1; -;
DR InterPro: IPR000606; -;
DR InterPro: IPR001788; -;
DR InterPro: IPR002588; -;
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183542 MW; 5DBBFB2FADCC5C0C CRC64;

Query Match 3.6%; Score 321.5; DB 1; Length 1616;
Best Local Similarity 19.4%; Pred. No. 1.7e-12;
Matches 352; Conservative 218; Mismatches 649; Indels 591; Gaps 83;

QY 197 TSYLLIHDDRAVYIEGDTSAQYHNHDSILRAMITTTIVGDHPIYERVAIGCHEVL 256
DB 247 INACFORDDRLTFSPASESTLNYSHSYSNILKYCKT----- 284
QY 257 LITAAPEPMPVPVPRSTEVYVRSIFPGGSPSLFPAACSTKSTFHAHPVHIMRLM 316
DB 285 -----YFP-ASREVIYMKF-----FLYTRVNTW----- 306
QY 317 FGATLDDQAFCC-CSRLMTYL--RGISKYVTGALVANEGMNSDALTXITAAYLTIGH 373
DB 307 -----FCKFSRIDFTLLKGVANH-----GVDEQYFKAMEDAMHYKKT---LAMCN 350
QY 374 -QRYLRTQALSKG-----MRRLGEHAQKFTIRLYWLEKSGR-----DIYPGRO 418
DB 351 SERILLEDDSSVYWPKKKDM-----VIYPLFISLETSKRYREVLVLKDFV----- 399
QY 419 LQFYAOCR--RMLSAGFHLDPRLVDFDESVPCHCFELKKVAGKFCFCFMRMLGQECTFL 476
DB 400 ---YTLNHIKRTYQAKALYTSNLSFEVSI--RSRYIINGVTAR----- 438
QY 477 EPAEGLVGDHNDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNPQDIARA 536
DB 439 -----SEMDVDRSLQSLSMTEFFLH----- 458
QY 537 SRLTATVEL-VASPDRLCETRYVGNKTFRTYVDGAH----- 572
DB 459 -----IKLAVLKDDLILISKFALPKPTVSOHWDEISLAFGNAFPIKERLINRKLKIT 512
QY 573 ---LEANGPEQYVLSFD--ASRQSMGAGSHL-----TYELTPA--GLQYRISNG 617
DB 513 ENALETRVPLVYTFHDLRYSEYKMSVDMBYLDIRKMEETEMVYALSELVYKKSDF 572
QY 618 DCTATPPPGAGSAPABGEVAFAFCALYRNRFTQRSLTGMLHDEGLGIFPPSPGH 677
DB 573 DVDV-----FSQMQSL----- 584
QY 678 IMESANPFCGEGTLTYRTMSTSGFSSDFSPPEAAAMAATPCLPSTPVSQ--TWLP 735
DB 585 ---EVPMTAAKVIYAVMSNESGLTLTFEQPTANVALA-----ODSEKASGALVYVS 636
QY 736 PSEEPQVAAAPVPAPDPDAGPGPVLPVLPVPPVHKSIIPPSNRRLYPPGAVY 795
DB 637 RDVEEESINGSMARGLQLAGLSDV-----PES-----SYTSEEL 673
QY 796 YAGSLFESDCDMLVNASNPGHPRGGGLCHAFYORPEAFYPTETI----- 840
DB 674 ESLEQFH-----MATASSLIHQ-----MCSIVYIGPLKYQCMKRFIDSLVASSAANSNV 725
QY 841 -MREGIAAYTLPRP---IIHVAAPDYRVEQNPRLKLEAAVRETCSRGTAAVPLGSGTY 896
DB 726 KILKDTAAIDLETROKFGVLDVASKRMVLVPSAKNAMGCVETRHAKYHVA--LLEHDE 783
QY 897 QVYVSLFPAWERNHPRGDELILTEPRAANFEAKRAQVLTITTEPTARANT-ALEIDA 955
DB 784 GI---ITCDWRR-----VAVSPESVYSOMAKRTLRLLKKG 819
QY 956 ATEVGRACAGCTISPGIVHGYFTAGVPGSGKRSIQG--DVDVAVVYPTRELNSWRR 1012
DB 820 EHVYSSA-----KVLVLDGVPGCGKTKELLSRVNFEEDLLIVPQROAAEMIRRR 868
QY 1013 ---GFAFTPTITANV-----TIG-----RRVVIDEAPSLPPLLIIH----- 1047
DB 869 ANASGIIVATKDVVRVDSFLMNYGARGAQFRFLFIDEG-----LMILHTGCNVFLVE 921
QY 1048 MRASSVHLLGDNPQIPALDFEHAQLVPA--IPELAPISWMVYTHRCAPDVCELLRGAY 1105
DB 922 MSLSDIAIYIGDQQLPYINRYTGFPYPAPHAFLVEDENETKRTTLRCPADVTHFINOR 981
QY 1106 P-KIQTTSRYLRSLFW-----NEPAIGQLVYVTOAKAA-----NPGATVYHE 1147
DB 982 EGHVMTSSSKRSVQSEWMSGASINPVSKPLGKILTLTFQSDKALLSGYDVTYHVE 1041

```

QY 1148 AGAATTTTII-ATADAGLIQSSRAHAIVAFRTETK-----CVID-ABGLLEEV-GI 1200
DB 1042 VGGTAYADVASIVRLTPPVSIARDSPHVLSSRHTSKSLXYTVVMDPLVSIIDLERV 1101
QY 1201 SDVIYNNFEL-AGGEVGHRSVYIPRGNPDONLGLQAFPPSCQSAVHOIAEELGHPA 1259
DB 1102 SSYLLDMKKNVAGYQXOLQVDVVF-----KNFLELVATPPTGDIS----- 1141
QY 1260 PVAAPVLPCEPELEOGLAY-----MPELTVSDDS--VLVEFLTDI-----VHCRRMAP 1304
DB 1142 -----DMQFYDKCLPGNSTLLNNYDAVTMLTDLISLVNKCCLIDMSKSVAP 1189
QY 1305 SQRKAVSTLVGRYGRTKLYEAHSDVRESLARIPT--IGVQATTCELYELVEANVE 1362
DB 1190 KDAFPTLIPMV-----RTAAEMPRGTGLENLVMIKRNFPSPELSGVDIENTASLVVD 1244
QY 1363 KGOGSANVELDLCNRDVSRTTFQKXCNKF-TGETIAHGKVG----- 1405
DB 1245 KFDYSYLKERRKPKNKS--LFSRESLNRIAKQEOVTTIGOLADFDVDDLPAVDQYRHM 1302
QY 1406 -----OGISAMSKTFCALGPGMPFAIEKEILL----- 1434
DB 1303 IKAQPKOKLDLSIQTEYPAQDTIYHSHKINAFGLFSELTQLDLSIDSRFLFTRK 1362
QY 1435 -PPNI--FYGAYEESVFAAASGASCMVENDPSEPTQNNFSLGLECVVMECGMP 1491
DB 1363 TPACIEDEFFGG-----LDHVPMADVELDYSKDKQNEHFCAVEYEIMRRLGLE 1412
QY 1492 QMLIRLVHVSAMVILQAPKESLKG-----WKHSGEPGLLMTNMAAIIAHC 1542
DB 1413 DFLAEV-----WKQGRKTKTLKDYTAGITCLMYQKSSDVTTFIGNITIIASCLASM 1465
QY 1543 YEFDFRVAAPKGDSSVVL-----CSDYRQSRNAALIAAGCKL-KVDYRPIGLYAGVV 1597
DB 1466 LPMKELIKGACGDSLLYFPKGC-EYEDIOQANILMNFPAKLFKQY--GYFCGRYV 1521
QY 1598 A---PGLTLPDVVRFAGRLSE---KNMGPEPEAEQRLAVCPRLGLTVAAQCV--D 1649
DB 1522 IHHDGCIIVYDPLKLSIKLAKHIKDW---DHLEFRRLSNVAVESLNCAYTOLD 1577
QY 1650 VSVRYGVSP 1659
DB 1578 AVGEVHKTP 1587

RESULT 13
RRPO_TOMK1
ID 090178: STANDARD: PRT: 1616 AA.
AC 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tomato mosaic virus (strain Kazakh KI) (TOMV) (TMV strain KI).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; tobamovirus.
ON NCBL_TaxID=138311;
RX MEDLINE=20196905; PubMed=10732356;
RA Belenovich E.V., Novikov V.K., Zavrlev S.K.;
RT "Biological properties and genome structure of the Kazakh isolate KI
RT of tobacco mosaic virus.";
RL Mol. Biol. (Mosk) 34:172-176(2000).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AJ243571; CAB62911.1;
DR EMBL; AJ243571; CAB62912.1;
DR InterPro: IPR000606;
DR InterPro: IPR001788;
DR InterPro: IPR002588;
DR Pfam; PF00978; RNA_dep_RNApol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
KM Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616
FT NP_BIND 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183606 MW; 4BF9913CE791DE33 CRC64;

```

Query Match 3.64; Score 320.5; DB 1; Length 1616;

Best Local Similarity 19.34; Pred. No. 2e-12; Mismatches 646; Indels 591; Gaps 83;

Matches 350; Conservative 223; Mismatches 646; Indels 591; Gaps 83;

```

QY 91 PNVLRGFLRPVGR--DVQRWYSAPTRG--PAANCRRSAL-RGLPPADRTYCFDGFSCA 145
DB 128 PNLVDVDMRHEGQKDSIELYLSRLERGNKHVPNFGQEAQFDRVAYEMNEVYCHDTFCR 187
QY 146 PAAE--TG-----VALYSLHDLMPAD-VAEAMARIGRTRVLAALLPREVLLPQ--TYHT 196
DB 188 HSQCYTGRYVALHLSIYDI-PADDFGAALRNVAHVCAAFHSENLLEDHSHVNLDE 246
QY 197 TSYLLHIDGBRAVYTYEGDTSAGYNHDVSLIRAMITTKIYGDHPLYIERVRAIGHFVL 256
DB 247 INACQFQDGRDLTFSPASESTLNSHGSYNTLKVCYT----- 284
QY 257 LITAPESPMPVYPYRSTREYVRSIFPGGSPSLFSPACSTKSTHAPVPHIMRLML 316
DB 285 -----YPP-ASNREVMKE-----FLVTRVNTW----- 306
QY 317 FGATLDQAFQ-CGRMTYV--RGISYKVYGAIVANEG--WMASEDALTXITAYLTIC 372
DB 307 -----FCKESRIDTFLYKGAHKG-----VDNEQFKAMEDMHHYKKT--LAMC 349
QY 373 H-QRYLRTQAIKSG-----MRRLGVEHAQKFTRLYSMLFEKSGR-----DYIPGR 417
DB 350 NSERILLEDSSVAYWPKKMDM-----VIYPLFIDISLETSKRTREKVLVSKDEV--- 399
QY 418 QLOFYAOCR--RWLSAGFHLDPRLVYFDESYPCCRCPTFLKVKAGKFCCEMRMLQECTCF 475
DB 400 ---YTVLNMHRTYQAALTYSNLSFEVESI--RSRYIINVTAR----- 438
QY 476 LEPAGELVGHGHNENAESEVDPAAPHLDVSGTAYGVAHCHOLEALYRLNVPQDIAAR 535
DB 439 -----SEMDVDSLSLQSLSTMTFFLH----- 458
QY 536 ASRLTATVELVASPRLCEPTVLGNKTFRTVVDGALHLEANGPROVYLSPDASROSGAG 595
DB 459 -----TKLAVLKDDLLISNFALGPKTVSOHWMEISLAGCN-----APPSIKERL--- 503
QY 596 SHSITVELTPAGLOVRISNGLDCTATPPPGAGPAAAGVAAFCALYRYNRFTQRHSL 655
DB 504 INRKLITENALEIRVP---DLXYVF-----HDLRLVSEYKM 537
QY 656 TGGIAML-----HPGGLGIGTPPSPGHIMESANPFCGEGILYRTWSTSGSPSFSPEA 710
DB 538 SYDMPVDIRKKMETEMNALSELVLTSDKFFVD-----VFSQMOCSLEVDPTA 591
QY 711 AAPMAATPG-----LPHSTPPVSDIWLPLPPSEEPQVDAAPVPAPDAPGLGPVYLP 765
DB 592 AKYIVAAVMSNMSGTLTFEOPTEANVALALQDSEK-----ASD-----GALVYT- 635

```



```

OY 766 PPPPVHKKPSTPPSRNRLTYTPDGAKVAGSLFESDCMLVNASNPHRGGLCHA 825
| : : :
DB 636 ---SRVPEPSI-----RGSMAEGLQ-LAGLS-----GDVPS 665
OY 826 FYORPEAFPTPTETMRGLAAYTLTPRIH-----AVADYVEQNPRLREAYRET 879
| : : :
DB 666 SYTRSEELSELEQFHM-----ATASLLHKQMSIYVYGPPLAVQOMKNIDSLVASL 717
OY 880 CSR-----RGTAAYPL-----LGSGLYQVPVLSFSDAM--ERNRHPDELYL 919
| : : :
DB 718 SAASVNLVKILKDTAIDLETROKFGVLDAVSKFMLVPSAANKHAMGVETIARKYHAL 777
OY 920 TE-----PAANWEAKKPAQPVLTITEDPARKTANLLEIDAENVRACAGCTISREI 972
| : : :
DB 778 LEHDEFGIITCDMWRVAVSSSEVYSPMAKIRTLRLIK-DGEPHVSSA----- 826
OY 973 VHYOPTAGVPGSGKRSIOOG---DVDVYVVPTRRLNSMRRR---GFAETPTAARV 1025
| : : :
DB 827 -KVVLVDGVPCCGKTKELLSRVNFEEDLLVPGQAEMIRANASGLIVATKONRTV 885
OY 1036 ---TIG-----RRVVIDEAPSLPHLLH-----MORASSVHLGDPNP 1064
| : : :
DB 886 DSFLNMYKGAQRCQKRLFIDEG-----LMLHTGCVNFLVEMSLCDIAYVYGPQOIP 938
OY 1065 AIDFHAHLVRA--IRPELAPTSMMXVTHRCPADYCELRGAYP-KIOTTSVLSLEW- 1120
| : : :
DB 939 YINRVTFGPYPAHFAKLVEDEYETRRTLRCPADYTHFLNORYEGHVMTSSEKSVSOE 998
OY 1121 ---NEPAIGOKLVYTOAKAA---NPGATVHEAGATFETTTI-ATAD 1163
| : : :
DB 999 MWSGAASINPYSKPLAKGILFTTQSDKELLSRGYADVTHVEVGEYIADVSLVRLPT 1058
OY 1164 ARGLIQSSRAHAVALTRTEK---CVILDA-----PGLLE 1197
| : : :
DB 1059 PVSIIAROSPVLVLSLHSTKSLKYTFVMDPLVSIIRDLERVSSYLLDMYKVDAGTOXQ 1118
OY 1198 VGISVYIVN-NPELAGSEVGHKRPVIRGNPDQNLGTLQAPRPGCOISAHYQLAEELGH 1256
| : : :
DB 1119 LQVDSVFKNFNLFAVAPKTG---DISDQFYDXC----- 1150
OY 1257 RPAPVAVALPPCELEOGLYMPQELTVSDS---VLVEFLTDI-----VHGRMAP 1304
| : : :
DB 1151 ---LPGNSTLLNNDAYTMKLTIDSLMWKCIDLMSKSVAP 1189
OY 1305 SORNAVSTLVGRYGRKRLYEANSDVRESLARIP--IGPVQATTCELYELVEAMVE 1362
| : : :
DB 1190 KDVKPTLLIPV---RTAAEMPROTGLENLVAMIKRNFNSPELSCVVDIENTASLYVD 1244
OY 1363 KGODSAVLELDLCNRDRIREFPOKXCKNF-TTGETIAHGKVG----- 1405
| : : :
DB 1245 KFFDSYLLKEKKPKNKS--LFSRESLNRMLAKQOQVITIGLADDFVYDLPAVDQYRHM 1302
OY 1406 ---QGISAMSKTFCALFGPWFPAIKEITALL----- 1434
| : : :
DB 1303 IKAQPKOKLDLSIQTEYRALQITIVHKKINAFGLFSELRLQDLSDSRFLFETRK 1362
OY 1435 -PPNI--FYGDAYESVFAAASGASCMYFENDSEPDSTONNFSLEGEVAMEECGP 1491
| : : :
DB 1363 TPAQLEDEFGD-----LDSHVPMQVLELDVSKYDKSONEHCAYEIVEIMRWLGE 1412
OY 1492 OMLIRLHYHVSAMWLAQKESLKG---WKHSGSPGTLMTVNNMAIIAHC 1542
| : : :
DB 1413 DELAEV-----WROGHKRTLLKDTAGIKTCLMYORKSGGVYTFITGTVIIASGLAM 1465
OY 1543 YEERDFRAAFKDDSVVL---CSDYROSNAALIAACGLKL-KVDRPRLGLYAGVV 1597
| : : :
DB 1466 LPMKRLINGAFCGDSDSLYFPKGC--EYPDIOQAANLMMWFELKRYKQY---GYCGRIV 1521
OY 1598 A---PGLTLPVVRPAGLSE---KNMGSPERADQRLAVACDPLRGLTNAQCV--D 1649
| : : :
DB 1522 IHHDGRCIVYVYDPLKLSIKLGAHKIKW---DHLEEFRRSLCDVAESLNCAYYTOLOD 1577
OY 1650 VVSRYGVSP 1659

```

```

DB 1578 AVGEVHTAP 1587
| : : :
RESULT 14
RPO_CRMV
ID RPO_CRMV STANDARD: PRT: 1597 AA.
AC 066220-
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE DE METHYLTRANSFERASE/RNA POLYMERASE (EC 2.7.7.48) (182 KDA PROTEIN) [CONTAINS:
OC CHINESE RAPE MOSAIC VIRUS (CRMV) (OILSEED RAPE MOSAIC VIRUS).
OX NCBI_TaxID=42007;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=96197410; PubMed=8616237;
RA Aguilar I., Sanchez F., Martin-Martin A., Martinez-Herrera D.,
RA Ponz F.;
RT "Nucleotide sequence of Chinese rape mosaic virus (oilseed rape mosaic
RT virus), a crucifer tobamovirus infectious on Arabidopsis thaliana.";
RL Plant Mol. Biol. 30:191-197(1996)
CC - FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC - FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC - MISCELLANEOUS: REAPTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1103 AND GLN-1105.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: U30944; AAB60599.1;
DR InterPro: IPR000606;
DR InterPro: IPR001788;
DR InterPro: IPR002588;
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1597
FT NP_BIND 1 1103 METHYLTRANSFERASE/RNA POLYMERASE.
FT NP_BIND 823 830 ATP (POTENTIAL).
FT VARIANT 919 919 V->G.
FT VARIANT 1286 1286 H->Y.
SQ SEQUENCE 1597 AA; 181621 MW; 7256A908BD3308F2 CRC64;

Query Match 3.6%; Score 320; DB 1; Length 1597;
Best Local Similarity 22.6%; Pred. No. 2,1e-12;
Matches 189; Conservative 131; Mismatches 302; Indels 216; Gaps 40;
OY 980 GVPGSGKRSIOOG---DVDVYVVPTRRLNSMRRGFA-----FTPT 1021
| : : :
DB 823 GVPGSGKRTLELEKYNFSDVLVYPEKESKMITIRANAGITRADKQVNVYDSTLMP 882
OY 1022 AARVTIGRRVVIDEAPSLPHLLHLMORASSVHLG-----DPNQIPAI---DF 1068
| : : :
DB 883 PRRVF--KRLFIDEG-----LMLHTGCVNFLMLSHCDVAVYVYDQOIFICVAFN 933
OY 1069 EH---AGLVPAIRPELAPTSMMXVTHRCPADYCELRGAYP-KIOTTSVLSLEW- 1122
| : : :
DB 934 PYPAHFAKLVDKEDR-----RYTLRCPADYTYFLNKGYSVLCISSVERSAEYV 987
OY 1123 ---PAIGOKLVYTOAK---AANPGATVHEAGATFETTTI-ATADAR 1165
| : : :

```

```

Db 988 RGKALNPITPLEGKILFTTQADKFLDLKGYKDVNTVHEVOGETYEKTAIVRLATPPL 1047
Cc 1166 GLIOSSRAHAVALTRHTEKC-----VILD-APGLLEVG-ISDVIY----- 1205
Cc 1048 EIIISRAHPHVALTRHTECKYTYVLDPMVAVISLGLSLNLEMYVESCTQXQLO 1107
Cc 1206 -----NNEFLAGEVGNHR-----PSVI PRGNP-----DONLQTLAAPPSCQ 1243
Cc 1108 IDIVFKGTNLFVTPKSGDMRDQFYDDTLFP-GNSTILNEFDVAVTNLNDISLNVKDCR 1166
Cc 1244 I--SAYHQLAEELGHRPA-----PVAALP-PCPELEFGLLLM-POELIYSDSLVLELT 1294
Cc 1167 IDPSKSYQVKE--RVEYKPKLRTAEMPRAGLENLVAMKRNMAADLTGTIDIE 1223
Cc 1295 DIVHCRMAAPSORKAVLSTLVGRYGRRTKLYEAHSD-----VRESLARFI-----PTIGP 1345
Cc 1224 D-----TASIVVEKFMADVYVKEFSGTGMAMTRESFSRWLSKQSSSTVGQ 1269
Cc 1346 VQ-----ATTCELYELVEMVEKGGDSANVELDLGNDVSRITTFQXCKNFTTGERT 1398
Cc 1270 LADFNFDLPAVDEYKMKISOPKQ-----KLDLSID-----EYPALQF 1309
Cc 1399 IARGKVGGISAMSKTFCALFEPWFAIEKEILLALPPN--IFYGDAVEESV--FAAAYS 1454
Cc 1310 IYV-----HSKTNALFGRPMFSELTMLETIDTSFLFTRTKPTQIEEFESDLD 1360
Cc 1455 GAGSCWVFNDFSEFSTONNFSLGLECYVMEEGMPQMLIRYHLVRSAMILQAPKESL 1514
Cc 1361 SSGAMETLEDISKYKDSQNEHFCAYEYKIMKLGIDMDLAEV-----WROGHRKTTL 1413
Cc 1515 KGF-----WKKHSGEGTLLMNTVMMALIAHCEYERDRFVAFAFKDDSVYLCSDY 1565
Cc 1414 KDVTAGIKTGLWYQKSGDVTTFIGNTIIIAACLSMIPMDKYAKAFCCDDSLIYLP- 1471
Cc 1566 ROSRNAALITAGGL-----KLKVDYRPIGLYAG-VVVAAPGIGTLP--DVVRFAGRLEKN 1618
Cc 1472 -KGLDLPDIOGANTLTNFEAKLFRRKYGFCCGRYVIHHRGALVYIDPLKLSLCKCKH 1530
Cc 1619 WGPGRERAQRLAVCDLFLGLTGNVAQCV--DVVSRYGVSPGLVNLGLMQLTIAD 1674
Cc 1531 I-RDEVILFEELRSLCDVTSNLNCAVEFSQLEDAVAEVHRTAVAGARVYCSIIKIYISD 1587

RESULT 15
RPO_TMOB STANDARD; PRT; 1616 AA.
AC P90211; 083484;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (163 KDA PROTEIN) [CONTAINS:
  METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tobamovirus OB.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
NCBI_TaxID=31749;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93389450; PubMed=8376970;
RA Ikeda R., Watanabe E., Watanabe Y., Okada Y.;
RT "Nucleotide sequence of tobamovirus OB which can spread systemically
  in N gene tobacco."
RL J. Gen. Virol. 74:1939-1944(1993).
CC -I- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
  RNA REPLICATION.
CC -I- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
  CAPING AND AN RNA HELICASE.
CC -I- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
  BETWEEN CODONS FOR ALA-1115 AND GIN-1117.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way

```

```

Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb.ch/announce/
Cc or send an email to license@isb.ch).
Cc -----
Cc EMBL: D13438; BAA02700.1; -
Cc DR EMBL: D13438; BAA02701.1; -
Cc DR InterPro: IPR000606; -
Cc DR InterPro: IPR001788; -
Cc DR InterPro: IPR002588; -
Cc DR Pfam: PF00978; RNA_dep_RNApol2; 1.
Cc DR Pfam: PF01443; Viral_helicase1; 1.
Cc DR Pfam: PF01660; YmeH/Ytransf; 1.
Cc KW Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
Cc FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
Cc FT NP_BIND 832 839 METHYLTRANSFERASE/RNA HELICASE.
Cc FT NP_BIND 832 839 ATP (POTENTIAL).
Cc SQ SEQUENCE 1616 AA; 182997 MW; 14E6BD679AA46575 CRC64;

```

Query Match 3.5%; Score 319.5; DB 1; Length 1616;
 Best Local Similarity 18.1%; Pred. No. 2,3e-12;
 Matches 329; Conservative 218; Mismatches 580; Indels 693; Gaps 74;

```

Cc 132 ADRTTCFDGSSRC--AFAETGVALYSLHDM--PAD-VAEMARGRYLAALHP 185
Cc 174 SDEVYCNFTQCCESNRYSSGGRVYALSLSLYDIPADEGALLKRNVHTLYAAHFAC 233
Cc 186 EVLLPCTYH--TTSYLLHDDGRAVVTYEGDTSAGYNDHVSILRAMIRTT-----KIV 237
Cc 234 ELLESTVELPTIGGIFGSDGKINCFNENSTLANTSHSYSLKLYCKITYTPPASRFV 293
Cc 238 GDHPLEIERYRAIGCHFVLLTLTAPESPMPVYVPYPRSTEVYVRSIFGPGSPLSPSAC 297
Cc 294 YKKEFLTRVNTWCFKFKLDT-----YTLY-----RGVYHRCDDQOEFSAM 336
Cc 298 ST-----KSTPHAV-----PVHMDRLM-----LEGATLDDQ----- 324
Cc 337 EDAMHYKKTILAMNLSERIVLEHSSVYVWPPKMDVIYPLFVLSLETKRTRKEYIVSK 396
Cc 325 -----AFCCSRMTYLRGISYKVTGALVANGMA----- 355
Cc 397 DEYTYLVNHRITYOAKALITYNNVLSFEYSIRSRVITINGVYARSEMDYDKLLOSMATFF 456
Cc 356 -----SEDALTAAYITAYLTI-----CHORY-----LRTQATSKGRRL 389
Cc 457 LITKLWMLKDELVSFKTLSAKSVHEVMDIKRGGNMPPSLKESLRLKLLSGAEL 516
Cc 390 GVHAAKFTIRLYSMLEKSGROYTPGRQL-QTYAQRRLSGPHLDPRVLYVDESVP- 447
Cc 517 EIEVPMYVTFHDFVAEKYKASYEMPTIDISKDLSAESYLSALSEL-----VLENSKDF 572
Cc 448 -----CRCRFLKVAAGKFCFMRWLGQECTC--FLEPREGLVGDHNEAVE-- 494
Cc 573 DLEKFSRMALINVDNIAKIYAAV--LSNESCGLLPKEPTEGMAEMKSGEDEVL 630
Cc 495 --GSEVDPAPPAHLDVSGTYAAVHGQLEALYRALNVPODIAARASRLTATVELVASPDL 552
Cc 631 TIGSQDNT-----DLSKSMVYISGLPICGIASEL 661
Cc 553 ECRIVYGNKTFRTTYVDGAHLENGPEQYVLSFDASROSGAGASHITELTPTAGLOVR 612
Cc 662 SCTCFVRNE-----EINSLEEYHML--AAESVYSNMASIVYS--GPIQVOO 704
Cc 613 SSGNLCCTATFPFGGAPSAAGV-----AAPCALRYRVRFTQBSLSTGLLHDE 664
Cc 705 MONYVSLA-----ASLSATVSNLKLKLVKDSVGVFODLSKGVFVRRK--MWL-- 752
Cc 665 GLIGIPPSPGHWEISANPFCGEGTLTYRTWTSGFSSDFSPPEAADAAMATPLPLPS 724
Cc 753 -----IKPLKLNKSWGVQKFDGKCFLLSY-----HN 781
Cc 725 TPVVSID-VLPLPPSEFOVDAVPADPAGLPBPVYLTPPPPPVHKPSITPPSRNR 783

```

```

Db 782 ELPCDADMSKVAVSNESV----- 801
QY 784 KLLIYPPGAKYAGSLFESDCDMLVNASNPGHRGGGLCHAFYORFPFAFPTFERIME 843
Db 802 -----YSDMAKL----- 808
QY 844 GLAAYTLFRPILIHAVPDYRVEQNPKRLEAAYRETCSRGTAAVPLIGSGIYOVPVSL 903
Db 809 -----RVLKRSIGEMPIVS 823
QY 904 PDAMERNHRPGDELYLTEBPANMFANKPAOPVLTTEDTARTANLALEIDATEVGRAC 963
Db 824 -----SAKYTL-VD----- 831
QY 964 AGCTISPGIVHQFTAGVPGSGKSRSIQG---DVDVVVVPTRRELNSWRRRGFAFPTH 1020
Db 832 -----GVPGCGKTREILRRVNSEDLVLPGEKAAMIRKR--ANQSGN 873
QY 1021 TAAR-----VTIG-----RRVVIDAPSLPPLL--LLHMQRASSVHLGDP 1060
Db 874 IVANNQNKTVDSFLMNLGKPGVCOFKRLFVDEGLMHPGVFLVKLSLCNEAFVPGDT 933
QY 1061 NOIPAIDE-----EHAGLVAIRPELAPTSWVXYTHRCPADVCELIRGAYP-KIQTTS 1112
Db 934 QOLPIYINRYQNPFPQHFSKLVDEFERKRT-----TLRCPDVTHFLNQCYDGAVTTS 988
QY 1113 RVLRSLEW-----NEPAIGOKLVXTQAK-----AANPGATVHEAGATFTE 1155
Db 989 KTORSGLEVGVGAAMNPVTKPLGKIYTFQSDKLTMLSRGQDVPNTVHEIOGETYEE 1048
QY 1156 TTII-ATADARGLIOSRAHAIVALTRHTEKC-----VILD-APGLREVGISDVIYVNF 1208
Db 1049 VSLVRLTPPTPIIISRESHVVLGLTRHT-KCFKITYVLDPLVKLVRLD---ECVSNF 1103
QY 1209 FL-----AGEVEXHNPVYIPRGNPDNLGTLOAFPPSCQISAYHQLAE 1253
Db 1104 LLDVYMWDSVSAKXQLOVSGVLAENLFVQAPKSGDQD---LQFYDKCLPGNSVTLNE 1159
QY 1254 LGHRAPAVAVLPCEBEQGLLYMPQELTVSDSVLFE-----LTDIVHCR 1300
Db 1160 -----FDVATMNCSDI-----SLNVKDCVLDFSKSVPLPRDNTKVPPTVPI--R 1201
QY 1301 MAPSOR--KAVLSTLVG-----RYGRRTKLYEAAH 1329
Db 1202 TAAERPRSGLEENLVAMIKRNFNSPELSGYDMENTASVADRFDSYFLKDKLSGCSL 1261
QY 1330 SDV-----RESLARFI-----PTIGPVQATTCELYELVE-----AMVEKQDGSAY 1370
Db 1262 GDSGCKNIIDRQALLRMMEKQEKSTIGLAD---YDFVDLPAIDQYRHIIKSPKQ-- 1314
QY 1371 LEILDICNRDVSKITPFOCKCNKFTTGETTAHGKVGOGISAMSKTFCALFPMFRAIEKEI 1430
Db 1315 -KLDLSIQ-----SEYPSLQRTIVY-----HSKKINALFGPIFSELTROM 1352
QY 1431 LALLPN--IFYGDAYEESV--FAAASGASGOMFENDEFSDTONNPSGLCEVME 1486
Db 1353 LSAIDTSRLTFTTRKTPEDIEEFSDLAHQPMVEVLELDVSKDKSONEHCAVEYEIMK 1412
QY 1487 ECGMPQMLIRLYHLVNSANILOAPKESLGF-----WKHSGEPGTLMTNTVMNA 1537
Db 1413 RLGIDEFLEAV-----WKQGHRTKTLKDYTAGIKTCLMYOKRSGDVTTFIGNVTIAA 1465
QY 1538 ITAHCEFFDFRVAAFKGDSDVYLCSDIQSRRAAALLIAGCGL-----KLKV 1584
Db 1466 CMASMLPMEKVIKAACFGDSDLY-----LPKGCCELPTIOSCANIMMNEFAKL 1513
QY 1585 DYRPIGLYAGVVVA---PELGLTPDVVRFAPAGRLSEKNMGPRPARQLRLAVODFLRGLT 1641
Db 1514 FKRTGYFCGRYVIHHDRAIYVDPKLIISKIGAKHT-TDKEHLEEFRIISLADVSKSLN 1572
QY 1642 NVAQVCV--DVVSRYGVSP 1659
Db 1573 NCAYYAAQLDEAVREVHKTA 1592

```

Search completed: May 30, 2001, 16:14:36
Job time: 288 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:06:58 ; Search time 43.81 Seconds
(without alignments)
2663.582 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGTTTAEQALAAANSALA.....FTETIKPVLDLINSIORVE 1698

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7448	82.6	1693	1 MNWME	genome polypeptide
2	7433	82.5	1691	1 A44212	genome polypeptide
3	1368	15.2	290	2 C48547	nonstructural prot
4	763	8.5	152	2 B38196	probable RNA-dirc
5	391.5	4.3	1646	1 MNWMS2	186k protein - cuc
6	390.5	4.3	2115	2 S38480	nonstructural prot
7	374.5	4.2	2205	1 MNWVRN	nonstructural poly
8	372	4.1	141	2 A38196	non-structural hyp
9	349	3.9	1844	2 S01956	hypothetical prote
10	346	3.8	1985	2 S19151	hypothetical prote
11	324	3.6	1616	1 J02144	183k protein - tom
12	323	3.6	1615	1 MNWME	180k protein - tom
13	320	3.6	1597	2 S65053	genome polypeptide
14	314	3.5	1601	2 S48699	178k protein - pep
15	313	3.5	1611	1 MNWMPV	183k protein - pep
16	311	3.5	1748	1 J01555	genome polypeptide
17	310	3.4	1844	1 RRPPTM	genome polypeptide
18	305.5	3.4	1776	1 RRPPTM	genome polypeptide
19	296	3.3	1615	1 MNWME	183.3k protein - t
20	293.5	3.3	1839	1 RRPPEM	genome polypeptide
21	293	3.3	1874	1 J00533	genome polypeptide
22	260	2.9	1608	1 MNWGM	183k protein - tob
23	243.5	2.7	1707	2 S01865	genome polypeptide
24	243.5	2.7	1884	1 A45353	genome polypeptide
25	242.5	2.7	1885	2 J02183	hypothetical 216.5
26	227	2.5	1112	2 S49432	replicase 126k - o
27	226.5	2.5	1766	2 S03701	141k protein - pea
28	218	2.4	2512	1 MNWVS	nonstructural poly
29	211	2.3	2514	1 MNWV82	nonstructural poly

30	207	2.3	2514	1 MNWV82	nonstructural poly
31	202.5	2.2	2492	1 C44213	nonstructural poly
32	202	2.2	1729	2 A49282	fusion protein 1a/
33	201.5	2.2	1707	2 S77910	hypothetical prote
34	201.5	2.2	1718	1 J01734	genome polypeptide
35	200	2.2	2431	1 MNWVSF	nonstructural poly
36	194	2.2	1968	1 PNO093	genome polypeptide
37	184.5	2.0	3175	1 RRPVEV	hypothetical prote
38	184	2.0	1707	2 S77908	hypothetical prote
39	183	2.0	1116	2 S26358	nonstructural poly
40	177.5	2.0	2492	1 MNWVTD	nonstructural poly
41	176	2.0	1385	2 S34230	156k protein - pia
42	174.5	1.9	2492	1 A44213	nonstructural poly
43	172.5	1.9	2493	2 S26372	nonstructural poly
44	169.5	1.9	2493	2 S72349	nonstructural poly
45	167	1.9	810	1 P2WMBB	2a protein - broad

ALIGNMENTS

RESULT 1

MNWMHE
genome polypeptide - hepatitis E virus (strain Burma)
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: hepatitis E virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Jul-1999
C:Accession: A40778; A48547
R:Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, Virology 185, 120-131, 1991
A:Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length
A:Reference number: A40778; MUID:92024067
A:Accession: A40778
A:Molecule type: genomic RNA
A:Residues: 1-1693 <TAM>
A:Cross-references: GB:M73218; NID:9330023; PIDN:AAA45734.1; PID:9330024
R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kilm, J.P.; Luk, K.C.; Young, L.M.; Piatak, M.; F
Virus Genes 6, 173-185, 1992
A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e
A:Reference number: A48547; MUID:92271462
A:Accession: A48547
A:Molecule type: genomic RNA
A:Residues: 967-1693 <FRY>
A:Cross-references: GB:M32400; NID:9330021; PIDN:AAA03206.1; PID:9330022
A:Note: sequence extracted from NCBI backbone (NCBIN:104572, NCBI:P:104573)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; nonstructural protein; nucleotidyltransferase

Query Match 82.6%; Score 7448; DB 1; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY	1	PGTTTAEQALAAANSALANAVVPRPLSRVOTELLINMOPROLYPPEVIMNPIOR	60
DB	10	PGTTTAEQALAAANSALANAVVPRPLSRVOTELLINMOPROLYPPEVIMNPIOR	69
QY	61	VHNELEOYCARAGRCLEVGAFHRSINDPNVLRHCFPLPVGRDQVRYSAPTRGPAAN	120
DB	70	VHNELEOYCARAGRCLEVGAFHRSINDPNVLRHCFPLPVGRDQVRYSAPTRGPAAN	129
QY	121	CRSALRGPLPADRTYCFGFSKCAFAETGVALYSLHDMPADVAEAMARHXTLYAA	180
DB	130	CRSALRGPLPADRTYCFGFSKCAFAETGVALYSLHDMPADVAEAMARHXTLYAA	189
QY	181	LHPPEVLLPPEGTHYHTSYLLIHGDRAVYTYEGDSAGYNHDSVILRAMIRTKTVGDH	240
DB	190	LHPPEVLLPPEGTHYHTSYLLIHGDRAVYTYEGDSAGYNHDSVILRAMIRTKTVGDH	249
QY	241	PLVIEFVRALIGCHFVLLTAAPEPSPMPYVPYPRSTEVVRSIFGSGSPSLPSPACSTK	300
DB	250	PLVIEFVRALIGCHFVLLTAAPEPSPMPYVPYPRSTEVVRSIFGSGSPSLPSPACSTK	309

QY 301 STEHAIVHINDRLMFCATLDDOAFCCSRMTYLRGISYKVTGALVANEGBWNSADAL 360
 Db 310 STEHAIVHINDRLMFCATLDDOAFCCSRMTYLRGISYKVTGALVANEGBWNSADAL 369
 QY 361 TAXITAAVLTICHOXYLTQOAIKSMRSLVGHAKCFITRILYSWLFEEKSGDHYTGRLO 420
 Db 370 TAXITAAVLTICHOXYLTQOAIKSMRSLVGHAKCFITRILYSWLFEEKSGDHYTGRLO 429
 QY 421 FYAACRRMLVSGFHLDPRLVFEDESVPQRCRTFLKRVAGKRCCEMRWMOGECTCLPEPAE 480
 Db 430 FYAACRRMLVSGFHLDPRLVFEDESAPCHCRALRKALSKCCCEMRWMOGECTCLPEPAE 489
 QY 481 GLVGDHGDNABESSEVDPAEPANLDVSGTYAHQHOLEALYALANPDIAPASRLT 540
 Db 490 GAVGOGGDNABESSEVDPAEPASLDSISGYVPGTALQPLQYALDPAETVAVAGRLT 549
 QY 541 ATVELVASPDREKTYVGNKTFRTYVDGAHLEANGPEQVYVLSDAOSGSGSHSLT 600
 Db 550 ATVELVASPDREKTYVGNKTFRTYVDGAHLEANGPEQVYVLSDAOSGSGSHSLT 609
 QY 601 YELTPAGLQVRISNGLDCTATFPFGAPASAPGEVAFAFCALYRNFTORHSLTGLM 660
 Db 610 YASASAGLEVYVYAGADHRAVFAVPSRABGEVTAFCALYRNFTORHSLTGLM 669
 QY 661 LHPBGLGIFPPSPBGHAWESANPFCGEGTLYRTWS-TSGFSSDSESP-PPAADA 714
 Db 670 FHPBGLGIFPPSPBGHAWESANPFCGEGTLYRTWS-EVDAVSPARPDLGFMSEPSIS 729
 QY 715 MAATPGLPHSTPRVSDIMVLPPESEEPQVDAAPV-PPAPDPAGLP-GVVVLTPPPPPVH 772
 Db 730 RAATPPLT-----AALPPAPDPSPPSAPALAEPPASGATAG 766
 QY 773 KPSP-PPSRNRRLTYTPDGAKYVAGSLFESDCLVNAANPCHRGHGGCLHAFFORFP 831
 Db 767 APALTTHQARRRLLETPDGSKYFAGSLFESTGWLVAANVDRHRCGGLCHAFYQYXP 826
 QY 832 EAFPTTEIMEGLAATYLPRLPIIHAAPDYRVEQNPRLBAAYRETCSRGTAAVPL 891
 Db 827 ASFPAASVVMGDGAAYLTPRLPIIHAAPDYRVEQNPRLBAAYRETCSRGTAAVPL 886
 QY 892 GSGIYOVVSLSPDAMENHRPDELTYLTPAANWFENKRAOPVLTTEPTAPRANAL 951
 Db 887 GTGIYOVVSLSPDAMENHRPDELTYLTPAANWFENKRAOPVLTTEPTAPRANAL 946
 QY 952 EIDATEVGRACAGCTISPGIVHYOFTAGVPSGSKRSIOGDVYVVVPTRELNSWR 1011
 Db 947 EIDATEVGRACAGCTISPGIVHYOFTAGVPSGSKRSIOGDVYVVVPTRELNSWR 1006
 QY 1012 RGFAPPTHTAARTYIGRRVYIDEAPSLPHLLHLMORASSVHLGDPNCPAIDEFHA 1071
 Db 1007 RGFAPPTHTAARTYIGRRVYIDEAPSLPHLLHLMORASSVHLGDPNCPAIDEFHA 1066
 QY 1072 GLVPAIRPELAPTSMWXYTHRCPADVCELIRGAYPKIOTTSRVLSRLPWNPAIGOKLVX 1131
 Db 1067 GLVPAIRPELAPTSMWXYTHRCPADVCELIRGAYPKIOTTSRVLSRLPWNPAIGOKLVX 1126
 QY 1132 TOAKAANPGATVHEAOGATFETTTIATADANGLIQSSRAHAIVALTRHTEKCVILDA 1191
 Db 1127 TOAKAANPGATVHEAOGATFETTTIATADANGLIQSSRAHAIVALTRHTEKCVILDA 1186
 QY 1192 PGLIREGISVIVNNPFLAGCEVGHRRSVIPRGNPNONIGTIOAPPSCQISAHYOLA 1251
 Db 1187 PGLIREGISVIVNNPFLAGCEVGHRRSVIPRGNPNONIGTIOAPPSCQISAHYOLA 1246
 QY 1252 BELGHRAPVAAVLPPELEOGLLYMPQELTSDVSLVFLTDLIVHCRMAAPSORAVL 1311
 Db 1247 BELGHRAPVAAVLPPELEOGLLYMPQELTSDVSLVFLTDLIVHCRMAAPSORAVL 1306
 QY 1312 STLVGRRGRTKLYEAASDVRESLARFIPTIGVQATTCLELYELVEAMVEKGGDSAVL 1371
 Db 1307 STLVGRRGRTKLYEAASDVRESLARFIPTIGVQATTCLELYELVEAMVEKGGDSAVL 1366
 QY 1372 ELDCNNDVSVITTFQCKCKNFTTGETIAHGKVGOGISAMSKTFCALGFPWFRAIEKEIL 1431

Db 1367 ELDCNNDVSVITTFQCKCKNFTTGETIAHGKVGOGISAMSKTFCALGFPWFRAIEKEIL 1426
 QY 1432 ALBPNTIFGDAEESVFAAANVSGAGSCMFENPSESDSONNPSGLCEVVEECMP 1491
 Db 1427 ALBPNTIFGDAEESVFAAANVSGAGSCMFENPSESDSONNPSGLCEVVEECMP 1486
 QY 1492 QMLIRLYHVLVSAMLIQAPKESLKGFMKHSCEGTLLMNTVMNAALIAHCEFRDFEPA 1551
 Db 1487 QMLIRLYHVLVSAMLIQAPKESLKGFMKHSCEGTLLMNTVMNAALIAHCEFRDFEPA 1546
 QY 1552 AFKGDSDVYLCSDYROSNAALIAAGCLKLVDRPPLGLYAGVAVAPGLTLPDYVFA 1611
 Db 1547 AFKGDSDVYLCSDYROSNAALIAAGCLKLVDRPPLGLYAGVAVAPGLTLPDYVFA 1606
 QY 1612 GRLEKMWGPEPERAEOLRLAVCPDLRGTLNVAQVQVYVSRVYGVSPGLVHNLIGMLOT 1671
 Db 1607 GRLEKMWGPEPERAEOLRLAVCPDLRGTLNVAQVQVYVSRVYGVSPGLVHNLIGMLOT 1666
 QY 1672 IADKAPHTETIKPVLDTLSIIORVE 1698
 Db 1667 IADKAPHTETIKPVLDTLSIIORVE 1693

RESULT 2
 A44212
 genome polypeptide - hepatitis E virus (strain Mexico)
 N:Contigs: RNA-directed RNA polymerase (EC 2.7.7.48)
 C:Species: hepatitis E virus
 C>Date: 17-Feb-1994 #sequence, revision 17-Feb-1994 #text, change 23-Jul-1999
 C:Accession: A44212; B48547
 R:Huang, C.C.; Nguyen, D.; Fernandez, J.; Yun, K.Y.; Fry, K.E.; Bradley, D.W.; Tam, A.
 Virology 191, 550-558, 1992
 A>Title: Molecular cloning and sequencing of the Mexico isolate of hepatitis E virus
 A:Reference number: A44212; MUID:93079857
 A:Accession: A44212
 A:Molecule type: genomic RNA
 A:Residues: 1-1691 <HUA>
 A:Cross-references: GB:M74506; NID:g330017; PIDN:AAA45730.1; PID:g330018
 R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platek, M.; F.
 Virus Genes 6, 173-185, 1992
 A>Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e
 A:Reference number: A48547; MUID:92271462
 A:Accession: B48547
 A:Molecule type: genomic RNA
 A:Residues: 965-1691 <FRY>
 A>Note: sequence extracted from NCBI backbone (NCBIT:104576, NCBI:P.104578)
 C:Superfamily: hepatitis E virus nonstructural protein
 C:Keywords: ATP; GTP binding; nonstructural protein; nucleotidyltransferase; P-loop
 F:973-980/Region: nucleotide-binding motif A (P-loop)
 F:979/Binding site: ATP/GTP (Lys) #status predicted

Query Match 82.5% Score 7433; DB 1; Length 1691;
 Best Local Similarity 81.08; Pred. No. 0;
 Matches 1384; Conservative 121; Mismatches 166; Indels 38; Gaps 5;
 QY 1 PGITTAEOAALAAANALANAVVVPRLSVQREIILINLMOPQOLYFREVLMNHP 60
 Db 10 PGITTAEOAALAAANALANAVVVPRLSVQREIILINLMOPQOLYFREVLMNHP 69
 QY 61 VINHELOYCARARQCLEVGAHPRISINDNPNVLRHCFLRPVGDVORWYSAPTRGPAAN 120
 Db 70 VINHELOYCARARQCLEVGAHPRISINDNPNVLRHCFLRPVGDVORWYSAPTRGPAAN 129
 QY 121 CRRSALGLPRADDTYCFDFGSCAFAPETGVALYSLHDLMPADVAAMARHGCTRXYAA 180
 Db 130 CRRSALGLPRADDTYCFDFGSCAFAPETGVALYSLHDLMPADVAAMARHGCTRXYAA 189
 QY 181 LHPPEVLLPFGTYHTSYLLIHDGDAVYVTEGDSAGYNHDSIRAWIRTKKIYGDH 240
 Db 190 LHPPEVLLPFGTYHTSYLLIHDGDAVYVTEGDSAGYNHDSIRAWIRTKKIYGDH 249

QY	241	PLVIERVA	IC	CH	V	L	L	T	T	A	A	P	E	S	P	P	V	P	P	R	S	T	E	V	E	V	A	S	I	P	G	S	P	L	P	E	P	A	S	T	K	300												
Db	250	PLVIERV	A	N	G	I	C	H	V	L	L	T	T	A	A	P	E	S	P	P	V	P	P	R	S	T	E	V	E	V	A	S	I	P	G	S	P	L	P	E	P	A	C	K	309									
QY	301	STFHA	V	P	H	I	M	D	R	L	M	L	G	A	T	L	D	O	A	F	C	S	S	L	M	T	Y	L	E	I	G	I	S	Y	K	T	V	G	A	L	A	N	E	G	M	N	A	S	E	D	A	L	360	
Db	310	STFHA	V	P	H	I	M	D	R	L	M	L	G	A	T	L	D	O	A	F	C	S	S	L	M	T	Y	L	E	I	G	I	S	Y	K	T	V	G	A	L	A	N	E	G	M	N	A	T	E	D	A	L	369	
QY	361	TAXI	T	A	A	V	L	T	I	C	H	O	R	L	T	O	A	I	S	K	M	R	L	G	E	H	A	O	K	F	I	T	R	L	Y	S	M	L	E	S	G	R	D	I	P	G	R	O	L	420				
Db	370	TAXI	T	A	A	V	L	T	I	C	H	O	R	L	T	O	A	I	S	K	M	R	L	E	H	A	O	K	F	I	S	R	L	Y	S	M	L	E	S	G	R	D	I	P	G	R	O	L	429					
QY	421	FYA	O	C	R	R	L	S	G	F	L	D	R	V	L	P	E	S	V	P	C	R	T	F	L	K	V	A	A	K	P	C	C	F	M	R	L	G	O	E	T	C	L	E	P	A	E	480						
Db	430	FYA	O	C	R	R	L	S	G	F	L	D	R	V	L	P	E	S	V	P	C	R	T	F	L	K	V	A	A	K	P	C	C	F	M	R	L	G	O	E	S	C	F	L	O	A	E	489						
QY	481	GLV	D	H	G	H	D	N	A	E	G	S	E	V	D	A	E	P	A	B	A	H	L	D	S	G	T	V	A	H	O	E	A	L	R	A	L	M	P	O	D	I	A	A	S	E	L	T	540					
Db	480	GLV	D	H	G	H	D	N	A	E	G	S	E	V	D	A	E	P	A	B	A	H	L	D	S	G	T	V	A	H	O	E	A	L	R	A	L	M	P	O	D	I	A	A	S	E	L	T	549					
QY	541	ATV	E	L	V	A	S	P	D	R	L	E	C	R	T	V	L	G	N	K	T	E	R	T	V	D	G	A	H	L	E	A	N	G	P	O	Q	Y	L	S	P	D	A	S	R	O	S	G	A	S	H	L	T	600
Db	550	ATV	E	L	V	A	S	P	D	R	L	E	C	R	T	V	L	G	N	K	T	E	R	T	V	D	G	A	H	L	E	A	N	G	P	O	Q	Y	L	S	P	D	A	S	R	O	S	G	A	S	H	L	T	609
QY	601	YEL	P	R	A	G	I	O	V	A	R	I	S	S	G	I	D	C	A	T	A	T	E	P	R	G	A	B	A	G	E	V	A	A	C	S	A	L	R	N	P	T	O	H	S	L	T	G	L	M	660			
Db	610	YAA	V	D	G	E	L	E	V	H	F	S	T	A	G	L	E	S	V	E	R	P	P	G	A	P	A	P	E	V	T	A	C	S	A	L	R	N	R	O	S	O	S	I	O	V	I	G	L	M	669			
QY	661	LHP	E	L	G	I	L	I	P	P	F	S	G	H	I	M	S	A	N	P	F	C	O	E	G	L	T	R	Y	T	W	S	T	-----	S	G	F	S	P	S	P	E	709											
Db	670	LHP	E	L	G	I	L	I	P	P	F	S	G	H	I	M	S	A																																				

```

Db      1303 VLSLTVGKRYGRIKRLDAGHTHDVRASLARIPTLGLKRVYATFTTELELVEANMEKQDSDSA 13623
QY      1370 VLEIDLNCNRDVSRLTEFFQKCNKFFTTGETIAHGKVQOGISAMSKFFCALFGWFAIEME 14299
Db      1363 VLEIDLCSRDVSRLLTEFFQKCNKFFTTGETIAHGKVQOGIFRNSKIFPCALFGWFAIEMKA 14222
QY      1430 ILALLPPIIEFYGDAYEEESVEFAAASGAGSCWVEENDESEFDSTONNFSLGLECYVMEEGC 14899
Db      1423 ILSLLPQAVFYFGDAVYDDSVSAAVAGASHAMTVEENDESEFDSTONNFSLGLECALMEEGC 14822
QY      1490 MPQWLIRLYHLVRSAMTLLQAPKESLKGFWKKHSGEPGLTMMVTVMNMAIIAHCEYFRDPR 15499
Db      1483 MPQWLIRLYHLVRSAMTLLQAPKESLKGFWKKHSGEPGLTMMVTVMNMAIIAHCEYFRDQ 15422
QY      1550 VAAEFGDDSVYLCSDYQKSRNMAALLAGCGLKTKLVDRPIGLYAGVVAAPGSLTLPDVR 16099
Db      1543 VAAEFGDDSVYLCSEYRPGASGLTAGCGLKTKLADPRPIGLYAGVVAAPGSLTLPDVR 16022
QY      1610 FAGRLSEKNMGPGGEERAEQRLAVACDFLKGTLTVAOVCYDVYSRYGVYSPGLVYHNLIGML 16699
Db      1603 FAGRLSEKNMGPGGEERAEQRLAVACDFLKGTLTVAOVCAICVEVYSPGVYSPGLVYHNLIGML 16622
QY      1670 QTADGKAHFTETIKPVLDTNSTIQARVE 1698
Db      1663 QTIDGKAHFTESYKPIILDTNSTIMHSE 1691

```

RESULT 3
nonstructural protein - hepatitis E virus (strain Tashkent) (fragment)
C48547
C:Species: hepatitis E virus
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 20-Sep-1999
C:Accession: C48547
R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platack, M.; F
virus Genes 6, 173-185, 1992
A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region
A:Reference number: A48547; MUID:92271462
A:Accession: C48547
A:Molecule type: genomic RNA
A:Residues: 1-290 <FR>
A:Cross-references: GB:I10337; NID:q291457; PION:AAA4573.1; PID:q291458
A:Note: Sequence extracted from NCHI backbone (NCBIN:104577, NCBI:P:104580)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; nonstructural protein

Query Match	15.2%	Score 1368:	DB 2:	Length 290:
Best Local Similarity	85.9%	Pred. No. 1.8e-76:		
Matches 249:	Conservative	21:	Mismatches 20:	Indels 0:
				Gaps 0:
QY 1344	GPVQATTCGLYELVEAMVEKGGDGSATVLELDLCNDRVSRTFFQKXCNFTTGETIAHGK	1403		
Db 1	GPVQVTTTCGLYELVEAMVEKGGDGSATVLELDLCNDRVSRTFFQKCNKFTTGETIAHGK	60		
QY 1404	VQGGISAMSKTTCALFGPMPFRAIEKEITLALPPNIFYGDAYEESVFAAASGASCMWYE	1463		
Db 61	VQGGISAMSKTTCALFGPMPFRAIEKAILLALPQGFYGAFPDVTFSAVVAAKASMWGE	120		
QY 1464	NDFSEFDSQNNFSLGLECVMECECGMPQWLRLYLVLVSAMTILQAPKESLGFPMKRHSG	1523		
Db 121	NDFSEFDSQNNFSLGLECEAIMEKCGMPKWLRLVHLILSAMIILQAPKESLGCCKRHSG	180		
QY 1524	EPGTLIMNTVMNAIITAHCEYEFRRFVAAFGKDDSVYLCSDYRQSIINAAALTAGCGLKIK	1583		
Db 181	EPGTLIMNTVMNAIITAHCHYDFRDQVAFAFKDDSVYLCSEKROSAGAVILLIAGCGLKIK	240		
QY 1584	VYRPPLGAYAGVVAAPGLCTLPDYYVRFPAAGRLSEKMNMGPEPERAEOLRLAV	1633		
Db 241	VGFRLPLGAYAGVVAAPGLALPDYVRLSGRLTEKNMGPEPERAEOLRLAV	290		

QY 1234 TLQAFPPSCQISAYHQIAEELGHRPAVAVLPPCELEBGLLYMPQELTVSDSVLVFEEL 1293
DB 1134 TM-----FATVPTKQLMQLNSLY-----1151
QY 1294 TDIHCRMAAPSGKRAVLSTLVGRGR-----RTKLYEAHSDVRESL 1336
DB 1152 ---VHNINFLPYSKGTGYTDMQEFYDRCLPGNSFVLNDEDAVTMRJRDHFNQLQPCRLTL 1208
QY 1337 ARTPT-----TQVQATTC-----LYELVEMAEKGODGSAYL-ELDLCNR 1378
DB 1209 SNLDPVALIKMEQNFLLPVLTACERPRIPQLLENLMAIKRNNTPDLAGTVDTITNM 1268
QY 1379 DVS-----RITPEOKCKNKFTGETIANGKVG-----1405
DB 1269 SISIVNFFSSFPVRDEVLDDHLCVRASSIQSFSDWFSQCPPTSAVQQLANFNFDLPARD 1328
QY 1406 -----OGISAMSKTFCALFGPWRFAIEKEILLAPRN--I 1438
DB 1329 TYMHMIRKQPKSRDLDTISQSEYPALQTIYVHPKVNAVAVGPPVKYLLTKFLSMVDSKFF 1388
QY 1439 FYGDAAVEEV--FAAIVSGAGSCMVFNDEPFSDTQNNFSLGECVMEECGMQMLIR 1496
DB 1389 FTRKKRPEDLOEFFSDLSHSDYELLDVSKYDQSQDSFHSIEMAIWEKGLDIDILAM 1448
QY 1497 LYHLVRSAMTLOAPKESLKG--FMKHSGBPGLTLMNT-----VNMMAIIAHCYEERDE 1548
DB 1449 MMSMGKRRITLDDFQAGITTLITYOKSGDYTFFIGNTFLIACVASMPLDKCFK-----1504
QY 1549 RAAAFKGGDSVVLCS--DYROSRNAALIAAGGLKLYKDYRPIGLYAGVV--APGILG 1602
DB 1505 --ASFCDGDSLLYLPKGLEYPDIQATANLY--WNFEAKLFRRKKGVEGCGYIIIHANGCI 1560
QY 1603 TLPDVYRFRGRSEKMGWGPPEAEQRLAVCDFGLGLINVA 1644
DB 1561 VYPPDPLKLISKLGKSL-VGEYHEVEFRISLIDVAHSLFNGA 1601

RESULT 6
S38480
nonstructural protein - rubella virus
C:Species: rubella virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C:Accession: S38480
R:Gillam, S.
submitted to the EMBL Data Library, March 1993
A:Description: Nucleotide sequence of the nonstructural protein genes of rubella virus.
A:Reference number: S38480
A:Accession: S38480
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-2115 <GIL>
A:Cross-References: EMBL:X72393; NID:g410507; PIDN:CAAS1087.1; PID:g410508
C:Superfamily: rubella virus nonstructural polyprotein

Query Match 4.3%; Score 390.5; DB 2; Length 2115;
Best Local Similarity 19.3%; Pred. No. 5.6e-16;
Matches 452; Conservative 205; Mismatches 736; Indels 947; Gaps 105;

QY 3 ITTAEOALAAANSAANLVVVRPLSRVQTEILINLMQRPQLVRRPREVLMNHPYQRYI 62
DB 40 VYTAQKRAIV-----AVIPRPVFTOMYS-----DHP--AL 69
QY 63 HNELEQYCARAGRCLEVAHPRS-----INDPNVLIHRCFLRPVGRDYQRMYS-----111
DB 70 H-AISYTRR--HWIEMG--PKALHLVLIDPSGL-----LREVARVERRWALCLNHT 118
QY 112 -----APTGRPAN-----CRSALRGIRP-----ADRYTC-FDGF 141
DB 119 ARKLATALLETSEAWHADYVC--ALRGAPSGPFYVHPEDVPHGGRVAVADRCLLYTTPM 175
QY 142 SRCFAAETGVALYSLHDLMPADVA-----E 167

DB 176 QMCELMRTIDATLLVAVDLMAVYALAAHVGDMDMDLGIAMHLHDGCGPADCGAGAGPTP 235
QY 168 AMARHGXTPLYAALHLPREVLLPRGYHTTSTYLLINDGRAVYTYEGDTSAGYNHDVSL 227
DB 236 GYTRPCTTRIYOVL--PDTAFHGRILYRCGPRLMWTFDCAVLAELSWEAONCHQ-----286
QY 228 RAMIRFTKIVGDHPLVTEVRAIGCHFVLLTAAPESPMPVYPRSTREYVYRSIFGPG 287
DB 287 -----ANRAVACTLPIRHVRSLOPS-----ARV-----311
QY 288 GSPSLFPASCTKSPFHAAPVHIMRLMFGATLDDQAFCCSLMTYLKGISYKVTGAL 347
DB 312 -LPDLVHLAEVGMWRNFSLPFPRYFQRMLSYCTLSDAIYSEVFEFKNALSHSTILAON 370
QY 348 VANEGN--ASEDALTYITAAVLTICHQRYLRTQAIKGMIRGLVEHQAOKFYRLYSW 404
DB 371 VLOEGWKGTCAEDALCAVAAR--AAQSNARLAGIMKSAKRCADSLT-----VAGM 421
QY 405 LPEKSGROYIPGRQLQFYA-----QRRMLSGFHLDPRLVLFDESVPCRCRFTLKYVA 458
DB 422 L-----DTIWDALIKRFSGSVPLAERMEK-----EQDAVAADF-----455
QY 459 GKFCCEPMLGQECCTCLEPABGLVGDHGDNEAYGESEVDAEPANLDSGTAYVHGQ 518
DB 456 -----RGPLEDGGRIHLDTVQ-----PKSPRPREIAATWYIHAAS 490
QY 519 LE--ALYALNVPOD-----IARASRLT 540
DB 491 ADRHCACAPRCQVPRERPSAPAGPPDEALIPWLEAERRALRCRMEDEALARADTAA 550
QY 541 ATVELVASPDRELCRYL-----GNKTRFTYVVDGAHLEANGPEQYVLS 584
DB 551 APAPLPPRAPRY--TVLRHRAHNGRMYLTDEBGGADALVLCDLQORLPKEPH---605
QY 585 FPAOSRMSGASHSLTYELTPAQIQ--VRISNGLDCTATFPFGGAPSAPEVAAFCSA 642
DB 606 -----YAAGAMCA--QARGLOAFVRY-----VPPERPMADGG--ARAMAK 643
QY 643 LYRYNRFOTRHSILTGLMLHPBGLG---IPEPSPGHMESANPCGEGTLTYRTWST 698
DB 644 FFRGCMAQR-----LTGEPAMHLPYTDGVDPKLI-----ALALRTLAQ 683
QY 699 SGFSSDPS-----PPEAAPAMATPG-LPHSTPVSYIWLPP-----SEEF 741
DB 684 QGALALASVRLDPRGTAFANAAVTAARAGPGLAATSPPGD---PPRRARRSQRH 739
QY 742 QVDAVAVPPAP--DPAGLCPGVVLTTPPPPVYKPSIAPP-----SNRRLLTYV-PD 791
DB 740 SDARGTPPRAPVRDP--RQF--SPAPRRVQDP--VPPTTAEPRARNAHAELEVYERS 794
QY 792 GAKVYAGSLFEED-----CDMLVNASNPGHRRPGGLCHAF 826
DB 795 GPPTSTKADPDSDIVESYARAAGPVHLRVLDIMDPKCKVYVNAANEGLLAGSGVCAI 854
QY 827 YQFPFAFYPTERIMEGLA--YTLTPRP-----ITNAVAP-----DY 863
DB 855 FANATAA-----LAADCRLAPCPIGEAVAATPGHGGCYTHIIHAUVAARRRDRDA 903
QY 864 RYEQNPKRLEAAVRE--TCSRGTAAVPLLGSITY--QVPSUSPFAMERNHPRGDE 916
DB 904 ALEEGALLERAVRSIVALLAARARARVACPLIGAGVYGMSAESIRALAAATRAAPER 963
QY 917 LYL-----TEBAANFEANKPAOPVLTITTEDTART 946
DB 964 VSLHICHPRATLTNHSVVGAGIARAVSPPTPELAS--CPAGDGRGROARSASPPAP 1022
QY 947 ANALEIDATEVGRACCTIS-----PGIV-----973
DB 1023 LG-----DATAPPRGCGCELCRYTRVTNDRAVYVNLERDGAATSWAMRIPEVVYVGP 1077
QY 974 -----HKQF-----TAGVPGSGKSR-----988
DB 1078 EHLATHFPLNHSVLKPAFVYRPPRGKMGSDMRRCRMQGMPOVRCITPSNAHALCTGVP 1137

Db 62 GFTS-----EFSISRAATP-----TP-----AAFLPPADDPSPTL 93
 QY 758 RGVPLVLPPEPPVHKPSIP-PPSRRRLTYTPDGAKVYAGSLFES 803
 Db 94 SAPARGEPAPGATAPARAATHTQTAHRRLLFTYTPDGSKVAFGSLFES 140

RESULT 9
 S01956
 hypothetical protein, 195k - turnip yellow mosaic virus
 C:Species: turnip yellow mosaic virus, TYMV
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Jun-1999
 C:Accession: S01956
 R:Morch, M.D.; Boyer, J.C.; Haenfl, A.J.
 Nucleic Acids Res. 16, 6157-6173, 1988.
 A:Title: Overlapping open reading frames revealed by complete nucleotide sequencing of
 A:Reference number: S01955; MUID:88289359
 A:Accession: S01956
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-1844 <MOR>
 A:Cross-references: EMBL:X07441; NID:962222; PIDN:CAA30322.1; PID:962224
 C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

Query Match 3.9%; Score 349; DB 2; Length 1844;
 Best Local Similarity 20.6%; Pred. No. 1.8e-13;
 Matches 385; Conservative 203; Mismatches 649; Indels 634; Gaps 93;

QY 157 LHD-----LMPADVAEAMARH-GXRLVLAHLPPVLLPPG-----TYHTTSLIHD 204
 Db 141 MHDLMVYHPSOIMDLFRKNPLERLVSLVPPRAHLSDSFPKLYTTTTRHTLHY- 199
 QY 205 GDRAVVTEGPTSGAGYNHDSILRAVITTKI-VGDHPLVIERVPAIG-CHEVILLTAAP 262
 Db 200 -----VPEGHAGSYNOP-SDAHSMLKINSIRUGNHLVTLIESGCVHSLIQETP 252
 QY 263 EPSMPVYVPRSTEVYRSIFGP----- 286
 Db 253 PPDSIQAPPTLMTSDLRQEPRLDVSPFRIPALIELPQATLQGLRDLRPVAYN 312
 QY 287 -----GGSPFLPSACSTKSTFAVNVPHIMDLMLGATLDDAFCCSRIMT 333
 Db 313 ALFTYTRAVRLRTSDPAFVRMHSSKPDHVMNSMADNIQT-----ALLNV 361
 QY 334 YLR-GISYKTVG-----ALVANGMNSSEDALTAIXITAAYLTICHQRYLTQAIKGM 387
 Db 362 PLRNVVYHVLQSPASISLKLROMRRL--TATAVPLISFLTL-ORFL----- 408
 QY 388 RLGVENHAQ-KFITRLYSWLFKESGRDYIPGRLQFYAOCRRMLSGFILDPRVLVFESEV 446
 Db 409 PLPRLPAEYKSTITAFRRRLYRKKE-----PHNPLDY-----FHLQHRVRYHSAI 453
 QY 447 PCRCRTFLKVVAGKFCCEMRWLQECCTFLPABGLVGDHGHNEAVEGSEVDAEPANL 506
 Db 454 ----- 511
 QY 507 DVSQTYAVHGHLEALYALNVPODIARA-----SRLTAVELVASPDR--- 551
 Db 465 KLP-----HALQKAALLLRPISPLLTATPPFRSEQKSMPLNAELSLMKFALPQWASL 519
 QY 552 -----LECRVYLGKNTFTTVDGAMLEANGDEQVYLSFDSAROSGAGSHSLTYELTPAG 607
 Db 520 VILALSSSILHLKLFSPPTIOAOH-----DYTHRHILHESYSLOMERPT-- 564
 QY 608 LOVRISSENGL--DCTATPPGAPSAAPGEVAFCFSAALYKRNRTQHSULTGLGLMHP 664
 Db 565 LSIPTRTAFLPFTPTTSTAPDRSEASLP---PAFAST-----FVPR----- 603
 QY 665 GLIGIIFPP--FSFG--HIMESANPFCGECTLTTRWSTSGF-----S 702
 Db 604 -----PPPAASSPGADPPTTTAAAPPIEPT--ORTHQNSDLALESSTSTPEPPPIRS 655

QY 703 SDFSP-----DEAAPAMAAATPGL--PHSTPPVSDIWLPPSEB----- 740
 Db 656 PDMTPSAVLEPEELNSRRRRPPQLPATPDLERATPPPLSTPHDDPDSADPLMGSHLLH 715
 QY 741 FOYDAAPVPPAPDAGLPGVVLTPPPPPV----- 775
 Db 716 HSLPAPPTHPRLPSSQLPAPLPTNDPTAIGVLPPEELHPRRYPPENTATFLRLSLPSNH 775
 QY 776 IPPSRNRRLTYTPDGAKVYAGSLFESDCDWLVNA--SNCGHPRG-----GLCH-- 824
 Db 776 LPQFTLN-CLLSAVSDQTKVSEELHWSLQTLPLDQSLSNETWTGLSTELHTALHLY 834
 QY 825 -----AFYQRPPEAFPTPEFTMRGLAAYTLTPPIIHAAVDYBONPKRLBAAYRET 879
 Db 835 NFOATVYSDRGPILFSPDITKR-----IDITHTPGPHSFGP-KRLIGSQP-- 881
 QY 880 CSRGGTAYPLLG-----SGITQVYVLSFDMERNHRGDELY---LTERAANFEA 929
 Db 882 -SARGHPSDPLIRAMSEKYSVGN-LPFS-----EAMHPTSIHAKNLISNKNKGF 933
 QY 930 NKPAQPVLTITEDTARTANLALTEIDAATEVGRACAGCTISPG---IVHYQFTAGVSG 985
 Db 934 VLSLIDYSTGQRTGPTPKERTIQIDHLYDT-----NPKTTPVYHF--AGFAGCG 981
 QY 986 KSRSTQ-----GDYDVVYVPTRELNSWR---RGFAA--FTPHTAARTIGRRV 1032
 Db 982 KTYPIQQLKTKLTKFDRP-VSCPTELTEKWTAMELHSGSMRFNWESSILKSSRLIV 1040
 QY 1033 IDEAPSLPHILLHMQARSSVHL--LGD-----NOIPALDFHAGIYV 1075
 Db 1041 IDELYKMRGYLDLSDLPDLLEVLIIIDPLAGEYHNSQSSDNHRLPS--ETLRLP 1097
 QY 1076 AIRPELAPTSMWXYTHRCPADVCELI-----RGAPYKIOT-----TSYVLSLF 1119
 Db 1098 YI-----DMYCNW--SYRLPCIALRLFOIHSNMAQVIGSVSTHDSQVLTNSHASSLT 1151
 QY 1120 WNEPAIQKLVYTOAANAANGAITVHEAOGATFETTTIATADARGLIOSSRAHAIVL 1179
 Db 1152 FN--SLGYR-----SCTISSQGLTFCDAPIVLDVYTKWLS--ANGVLAL 1194
 QY 1180 TR----- 1210
 Db 1195 TRSSSGVQFMGPSSYVGGTNGSSAMFSDAFNNSLIMDRFPSSLFPL--KLITSPILT 1251
 QY 1211 AGEVGHKRP----- 1241
 Db 1252 RGPRLNGATBSASPTNHSPPNHLRPHILPSTDRQFVYVNPPLPDQGETRLDT--HFLPP 1309
 QY 1242 QOISAVHQAEEILGHRPAPVAVALPQCELEOGILLYMPBELTVSDVLYVEL-----T 1294
 Db 1310 SRLPLHFDLPRAI--TPPPYSTVDP--PQAKASVY--PGEF--FDLSIAFFLRPHNDSTR 1363
 QY 1295 DIVICRNAAPS---QKRAVLSLVGRYGRKTKYLEAH-----SDVRSLARFI 1340
 Db 1364 EILHKDSSNQFPWFDRPFLSC-----QPSLSISAKHAPNHDPPTLIPASINRKL-RFR 1416
 QY 1341 PTIGPVQATT-----C-----ELXELVEMAVEKGDSAVLE 1372
 Db 1417 PSDSPHOITADVVGLGLFHSILCRAYSRODNSTVPPNPELFAACISLNETYAQLSSKTOS 1476
 QY 1373 LDCNDRVS-----RITPEOKXCNFTTGETIANGKVGQISAWSKTFCALRGP--W 1422
 Db 1477 TIVANASRSDPDMRHTTVYKIFAKQKHVNGSIEGSKAKOQTLALMDHYLVILGVPKKY 1536
 QY 1423 FRAIEKEILLALPPNIF-----YGAIVESVFAAANSVAGASCWAFENDSEFSTQ 1473
 Db 1537 QRIDN--ADRPNIYSHGKTPNQLRDMQOEHL-----TSTPKRIANDYAFPOSO 1586
 QY 1474 NNFSLGLECYVMECGMPOMLIRLY-HLYVSAMWLQAPKESLAKFKMKHSGEPCTLMNT 1532
 Db 1587 HGESVLEALMKRLNIPSHLIQHLVHKTIVSTQFGPLTCM-----RLTGEPRGYDNT 1641

QY 1382 ---RTTFQKCKNFTTGETTAHCKVGOGISAMSKTCALFGR--WFRATEKITALLP 1435
 Db 1490 RHTTVKTFKRAQKHNDGSLFGSKACQTLALMDHYVILVGPVKKQRIJFDN---ADRP 1546
 QY 1436 PNIF-----YGDAYEESVFAAASGAGSCWVFENDESEPDSTONNPSLGEQVME 1486
 Db 1547 PNITSHGCKTPNQRLDMCOEHL-----THSTPKIANDYTAFOQSGESVLEALKMK 1599
 QY 1487 ECGMPQMLIRLY-HLVASAMILQAPKESLKGFWKHSGEPTLLMTVMNNAITAHCEYF 1545
 Db 1600 RLNIPIHLIOLHVLKTNVSTQFGPLTCM-----RLTGEPTDYDNDNDYNAVYSQYDV 1654
 QY 1546 RDEYVAAPKDDSVYLCSDY--ROSRNAALIIACGLKLVADY-----1586
 Db 1655 GSCPIWV-SGDDSLI---DHPLPTRHDPMSVLKRLHLRFKELTSHPLFCGYVVPAGCI 1710
 QY 1587 -RPIGLAGVAVVAPGLTLPD-----VYRF--AGRLSEKNMGPERAEQRLAVCPF 1636
 Db 1711 RNPLALCKLMIAYDDALDRLSLYTEFTTHLGBESLMLHPETHVYOVSACPDF 1768

RESULT 11

183K protein - tomato mosaic virus (strain Ob)
 N:Contains: 126K protein
 C:Species: tomato mosaic virus
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 16-Jun-2000
 C:Accession: JQ2144; JQ2143; JQ2158; JQ2157
 R:Padgett, H.S.; Beachy, R.N.
 Plant Cell 5, 577-586, 1993
 A:Title: Analysis of a tobacco mosaic virus strain capable of overcoming N gene-mediated
 A:Reference number: JQ2143; M0ID:93299124
 A:Molecule type: mRNA
 A:Residues: 1-1616 <PAD>
 A:Cross-references: GB:111665
 A:Note: this form is translated based on a read-through of the codon TAG for residue 111
 A:Accession: JQ2143
 A:Molecule type: mRNA
 A:Residues: 1-1115 <PA2>
 A:Cross-references: GB:111665
 A:Note: R. R. Watanabe, E. Watanabe, Y. Okada, Y.
 J. Gen. Virol. 74, 1939-1944, 1993
 A:Title: Nucleotide sequence of tobamovirus Ob which can spread systemically in N gene
 A:Reference number: JQ2157; M0ID:93389450
 A:Accession: JQ2158
 A:Molecule type: genomic RNA
 A:Residues: 1-152, 'K', 154-872, 'N', 874-1616 <IKR>
 A:Cross-references: GB:D13438; NID:9436229; PIDN:BAA02700.1; PID:9436231
 A:Note: the codon TAG for residue 1116 is translated to Trp
 A:Accession: JQ2157
 A:Molecule type: genomic RNA
 A:Residues: 1-152, 'K', 154-872, 'N', 874-1115 <IKR2>
 A:Cross-references: DDBJ:D13438; NID:9436229; PIDN:BAA02701.1; PID:9436230
 C:Comment: This protein is involved in replication of the RNA genome.
 F:Superfamily: cucumber mosaic virus RNA 1 protein
 F:1-1616/Product: 183K protein #status predicted <MAN1>
 F:1-1115/Product: 126K protein #status predicted <MAN2>

Query Match 3.6%; Score 324; DB 1; Length 1616;
 Best Local Similarity 18.0%; Pred. No. 5,6e-12;
 Matches 337; Conservative 221; Mismatches 603; Indels 714; Gaps 75;

QY 94 LHCFCF---LRPGRDYORWYSAPTRGPANCRSAL-----RGLPADRTYCFDG 140
 Db 123 VHCGRMLDLDRIMRHENQDSVATYISRLNARKKVLPAQOEAFORTSRSSEVVCNNT 182
 QY 141 FSRG---AFAAETGVALYSLADLM--PAD--VAEAMARHGXTRLYALHLPEVLLPGTY 194
 Db 183 FCCESNRYSGGGRVVAISLHSLYDIPADDELGAALLKKNVHTLYAARHFADELLLEVSTV 242

QY 195 H--TTSYLLHGDRAVYVEGDTSGAGYNDHVSILRAMIIRT-----KIVGHPVIER 246
 Db 243 ELPTIGIFISFGDGIKINFCESNESTLNSHSYMLKYCKTYEPASNRRFVYMEFLITR 302
 QY 247 VRAIGCFVLLTLTAPRPSMPYPPYPRSRSEVYVRSIFGGGSPSLPSPACST-----KST 302
 Db 303 VNTWFCFTKLDI-----YTLY-----RGYHHGCDQOEYSAMEDAMHYKKT 345
 QY 303 FHAV-----PVHMDRLM-----LEGATLDDO-----324
 Db 346 LAMLSERIVLEDDSSVNYNFPKMKDVIYPLFPVSLFETQKRTKEVIYSKDPYTYVLNH 405
 QY 325 -----AFCCSRIMTYLRGISYKVTVALVANGMA-----S 356
 Db 406 IRTYQAKALFYNNVLSEVESIRSHVINGYARSEMDVALLQSMAMTFEFLITKLSMLK 465
 QY 357 EDALTAITAYLTI-----CHQY-----LFTQAISSKMRRLGVEHAKFT 398
 Db 466 DELLVSFTLSASVHSHVDEIKRGGMNFPSLKESLIRKKLISGSAELETIEVPMYV 525
 QY 399 TRLYSMLEKSGRDYIPRQL-OFYACRRMLSAGFHLDPRLVYFDESVP-----C 448
 Db 526 TENDRFYAEYKASVEMPTIDISKLSAEASYALSLSL-----VLENSKDFLEKFSRMC 581
 QY 449 RCRTFLKRVAGKPCCFMRWLGOBCTC--FLPAPAGLVGDHNDNAYE---GSEVDP 501
 Db 582 AINCVDNDIAAKIVAV--LSNESGVTLPKPEPEGMAAMKSGEKDEVLTGSGQDNT 639
 QY 502 EPHALDVSGTYAAVHGHOLEALRYALNVPDIIARASLTITVELVASPDLECRTVIGNK 561
 Db 640 -----DLTSKSMVISGLPCLGIASISDTEYRNE 670
 QY 562 TFRFTVVDGAHLBANGPEQYVLSFDASROSMGAGSHSLTYELTPAGLOVRISNGLDCTA 621
 Db 671 -----ELNSLEEHML--AAESYISKMSIYVS---GPLYQOQOMONYDLSLA 713
 QY 622 TTPPGGAPSAAPGEV-----AAPCSALYRINRTQHSITGGLMLHPEGLGTFPPF 673
 Db 714 -----ASLSATVSNLKLKLVKSDSVGFQDSLSKGVDFVRKK---MWL-----IKPT 756
 QY 674 SPGHIMESANPFGCEGTLTYRTWSTSGFSSDFSPPEAAAMATPGLPHSTPVSQI-W 752
 Db 757 LKNHSMKVQKFGKCFLLALST-----HNEPLTCADW 790
 QY 733 VLPPSEEFQVDAPVAPPADPGLDPGVLTTPPPPVHKSIPPSRRRLLYTPDG 792
 Db 791 SKVAVSNESMV-----YSDM 805
 QY 793 AKVYAGSLFESDCDWLVNASNPGHREGGLCHAPYORPEAPYPTFEIMREGLAAYTLTP 852
 Db 806 AKL-----808
 QY 853 RPIIHAVPDRVEQNKRLAEAVRETCSRGAAYPVLGSGIYQVPSLSPAMERNHR 912
 Db 809 -----RVLKSIIGEMTISV-----823
 QY 913 PGDELYLTEPANWFPAKPAQVLTTEEDTARTANLALETIDATEVGRACACTISPGI 972
 Db 824 -----SAKVTYLV-----831
 QY 973 VHYQTAGVGSKRSIQG---DVDVYVVPYRELNSRRRGPAFTPHTAARYTIG- 1028
 Db 832 -----GVPGCKTKTEILRNVFSEDLVLPCKEAAAMIRKRA-----NOSGKIYANN 878
 QY 1029 -----RRVVIDEAPSLPPHLL--TLHMORASSVYHLGDDPNOIPA 1065
 Db 879 DNWKYVDSFLMNLGKGPVCOFKRLFYDEGLMLHPCGYFLVKLSLSCNEAFVSGDIOIP 938
 QY 1066 IDF-----EHAGLVPAIRBELAPTSWAVTHRCPADVCELIRGAVP-KIQTTSRYLNS 1117
 Db 939 INRVQNEPFPQHSKLIIVDETEKRT-----TLRCVDVYTHLNGCYDGAVTTTSYQNS 993
 QY 1118 LFW-----NEPAIGOKLYXTQAAK-----AANPGAITVHEDAGATFTETTTI- 1159

A:Cross-references: EMBL:229370; NID:6488713; PIDN:CAA82559.1; PID:9619908
 A:Experimental source: tobamovirus infecting cruciferae plants (cr-TW)
 A>Note: readthrough of the terminator UGA occurs between codons CAA for 1107-Gln and CAA
 A>Note: the internal stop codon is translated as X
 C:Superfamily: cucumber mosaic virus RNA 1 protein
 F:1-1601/Product: 178K protein status predicted <PRO2>
 F:1-1107/Product: 122K protein status predicted <PRO1>

Query Match 3.5%; Score 314; DB 2; Length 1601;
 Best Local Similarity 21.5%; Pred. No. 2.3e-11;
 Matches 204; Conservative 149; Mismatches 338; Indels 256; Gaps 48;

893 SGIVY-----PVSLSDFAN-----ERNHRGDELYLEPAAHFAKPAQPLVIT 940
 Db 736 SGWVDVRCGRWMLIKPNAKS--HAWGVAEDAHK-----LVIVLLN--DDCKP-----VC 781
 Qy 941 EDTARTANLALTEIDATEVEGRACAGCTISPG-----IVHYOPTAGVSGSKRSIQ 991
 Db 782 DEWFEVAVASDLSIYSDMKLKTITCSPNGEPPEPNKAVI---LVDCVPGCGKTKETI 838
 Qy 992 QG---DVVVVVVTRRLNSWRRRGFPA-----FTPHTAARTIGRRVVI 1033
 Db 839 EKVFSEDLIVPGKSKMIIRAHNAGYIRADKDNVSTVDSFLMHPSRVP--KRLFI 896
 Qy 1034 DEAPSLPHLLH-----MORASSVHLGDPNQIPAI---DF---EHAGLVPAI 1077
 Db 897 DEG-----LMHTGCNVFLLLSQCDVAVYVGDQIPIFCIRVAVPYPAHFAKLVAD 949
 Qy 1078 RPLAPATSMWXYTHRCRADVCCELIRGAYP-KIOTSRVLSLFWNE-----PA 1124
 Db 950 EKRYR-----RTLRCPADVTYFLNKKYDCAVAVCTSAVESVAAEVYRGKALNPITLPL 1004
 Qy 1125 IGGKLVYQAAK-----AANPGAIVHEAGATFETTTI--ATADARGLIOSSRAHAIVA 1178
 Db 1005 EGRILFTFOADKPELLEKGYKDVNVEVOGETEKTALVRLSTPLEITISRAHPYIVA 1064
 Qy 1179 LTHHTKCK---YILD-----AGGLREVCISIVYI--NPFELA 1211
 Db 1065 LTHHTTRCKYVTVLDPMNVNISEMEKLSNELLDMYRVENGIOXOLOIDAVFKGTMLFVO 1124
 Qy 1212 GGEVGYHR-----PSVIPGPN-----DONLGLQAFPPSCOI--SAVHOLAEL 1254
 Db 1125 TPKSGMRDMQFYNDTLPL--GNSITINEYDAVTMLRDISLANKDCRIDSKSVQLPKE- 1182
 Qy 1255 GHRPAPV-----AAVLPPCELEOGLLYM--POELTVSDSVLVELDVIHCRMAAPS 1305
 Db 1183 ---QPIFLKPKIRTAAMPRTAGLLENLVAMIKRMNNADPLGTIDIED----- 1228
 Qy 1306 QKRAVISTLVGRY--GRRTKLYEAH--SDVRESLARFI-----PTIGPVQ-----AT 1349
 Db 1229 ---TASLVYERKFWDSYIDKEFSGTNEMTWTRESFQWLSKQESSSTVQGLADENFVDLPA 1284
 Qy 1350 TCGLYELVEAMVAKGODGSVLELDLCNRDVSRTFFQCKCNKFTGETIHAHKVGGIS 1409
 Db 1285 VDEYKIMISQPKQ-----KLDLSIDP-----EYPLQITIVY----- 1316
 Qy 1410 AMSKTCALFGPWFAIEKEIALLRPN--IFYGDAVEESV--FAAASGAGSCWFEEND 1465
 Db 1317 -HSKKTINATJFGPMFSELTFRLLERIDSKFLPYTRKTPAQIEDFFSDLDSTQAMEILFD 1375
 Qy 1466 FSEFDSQNNFSLGLECVWEECGMPOMLRLYLHLVRSAMILDAKESLKF----- 1517
 Db 1376 ISKYDSQNFHCFAVEKIMKLGIDEMLAEV-----WKQGHRTKTLTYDAGIKYCTL 1428
 Qy 1518 -RKHSGEPECTLLMNTVYNNNAITAHCEYEFDFVAAAFKGGDSVVLQDSQSNAAALIA 1576
 Db 1429 -WYDRKSGDVTYTGNTIITACLSMIPMDKVIKAAFCGDSLIYIP--KGLDLPDICA 1485
 Qy 1577 GCGL---KLVYDRIPLGLVAG--VVVAPGLGTLPL--DVNFAGRLSKNNPGERAEOL 1629
 Db 1486 GANLMMNFKAELPRKKYVCGCGRYVIHHDGALIVYVDPLKLSKLGCKHI--RDVHLEEL 1544

Qy 1630 RLAVCDFELKGLTINVAQCV--DVYSRVYGSVPSGLVNLIGMLOTIAD 1674
 Db 1545 RESLQVNASMLNCAIFYSQLDEAVAEHVTAVGSGFAFCSITIKYSLD 1591

RESULT 15
 WMTMPV
 183K protein - pepper mild mottle virus (strain Spain)

M:contains: 126K protein
 C:Species: pepper mild mottle virus
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Apr-1994
 C:Accession: J01312
 R:Alonso, E.; Garcia-Lange, I.; de la Cruz, A.; Wicke, B.; Avila-Rincon, M.J.; Serra, J. Gen. Virol. 72, 2875-2884, 1991
 A:Title: Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resist
 A:Reference number: J01312; MOID:92113528
 A:Accession: J01312
 A:Molecule type: genomic RNA
 A:Residues: 1-1611 <AIO>
 A:Cross-references: GB:M81413
 A>Note: readthrough of the terminator UAG occurs between the codons CAA for 1117-Gln
 C:Comment: This protein may have RNA polymerase activity.
 C:Superfamily: cucumber mosaic virus RNA 1 protein
 F:1-1117/Product: 126K protein status predicted <PRO>

Query Match 3.5%; Score 313; DB 1; Length 1611;
 Best Local Similarity 18.7%; Pred. No. 2.7e-11;
 Matches 341; Conservative 227; Mismatches 642; Indels 610; Gaps 79;

94 LHRCELPVRVDQVRYSATRRPANCRSAL-----RGLPADRTYCPD--- 139
 Db 123 VHCMPNMDLRDVR-----HNAQDSLELYLSLAKOKKKVLPYQKP-CEQKYT 171
 Qy 140 -----GFSCAFAETGVALYSLHLDLPAD--VAEAMARHGXTRLYAALH 182
 Db 172 DDQSVYVCSKPFQHCBSVSHCTDKV--YAVALHSLYLD--PADDEGALLRRNVHVCYAAH 229
 Qy 183 LPPEVLLPCTYHTSYLLIHD-----GDRAVVTYEGDTSAGYNHDSVILRAMI--- 231
 Db 230 FSENLIL-----EDSVSILDDIGAFPSREGDMINSEFVAESTLNTYHSYSLNLYKVCYCKT 283
 Qy 232 ---RTTKYGDHDLVIERVRAIGCHFYLLLTAAPEPMPMYVYVPRSTEVYVSGPGG 288
 Db 284 YFPASSREYVMEKFLVTVRVTFCKFSRL-----DTEVLVYGVYHGV 326
 Qy 289 SPSLFSPACST---KSTF-----HAYVHIIMRLM-----LFGATLDDQ--- 324
 Db 327 DKQGFYSAMEDAHYKKTLLAMNSERTLLEDSSSVNWFPMKMDVYVLPFDVSLONEGK 386
 Qy 325 -----AFCCSRIMTYLRGISYKVTVGALVANEGWASE 357
 Db 387 RLARKEVYVSKDPVYTVLNIHRTYOSKALTYANVLSEVESIRSRVITINGVYASMDV-D 445
 Qy 358 DALTAHTAYLYLTICHRVYLRQAISGMRLRGEVNAQKITLILYSLEFKSGRDYIPGR 417
 Db 446 KALLQSLSMFELQTKLAMLKDLV-----VQKE-----QVHSKSLTEYV--- 485
 Qy 418 QLOFYAQCRW--LSAGFHLDPRVLVDESVPCRFTFLKAVGKFCFARMJGOECTCF 475
 Db 486 -----WDELTAFH-----NCPFTKE-----RLNKKLITY 512
 Qy 476 LEPA-EGLVGD-----HGHNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNP 529
 Db 513 SEKALEIKYVDLVYTFHDLRVKREKSSVEMPV-----LDVKKSL---EEAEVYNNAL--- 561
 Qy 530 QDIAARSRITAVEVLAASDRLECRVTYGNKTFRTTYVDGAIHLEANGPQOYVLSFASR 589
 Db 562 -----SEISLIKSDSKFD-----VDVESRMC 582
 Qy 590 QSMGAGSHSLYELTTPAGLOVRISNGLDCTATFPFGAGASAAGEVAACSAALYRYNR 649
 Db 583 NTLGVD-----PLVAAKVVAVAVVSNESGLTLIFE---RTEANVVALA----- 621

```

QY 650 TORHSLGGLMLHPEGLGIFPPSPGHIMESANFCEGTLTYRTWSTSGFSSDSPPE 709
Db 622 -----LQPTITSKEGSLKI-----VSSDVGESSLIKE 648
QY 710 AAAPMAATPCLPHSTPPVSDIWLPPSEEFQVDAAPVPPAPBAGLPGVVLTPPPPP 769
Db 649 VVRKSEISMGLTGNT--VSDEFORSTEIESIQ-----679
QY 770 PVHKPSIPPPRRNRRLTYTPDGAKVYAGSLFESDC--DMLVNASNPHRPGGGLCHA 825
Db 680 QFHVAVSTFTIRKQK-----HAMVYTGPLKVOQCKRYLDSLV-----716
QY 826 FYQRPPEAFYPTFEIMREGLAAYTLPPRIHAAVADYVEQNPRLKLEAAYRETCSSRGT 885
Db 717 -----ASLSAAVSNLKKIHKDPA--AIDLETKERKGVY-DVCLKKML 755
QY 886 A-----AYPLLSGIGYQVVPVSLFDMWERNHRPGDELITPERANWFENKRAQPVLT 938
Db 756 VKPLSKGHAMGVVMDSDKCFVAL-----LTYDGEN-----I 787
QY 939 ITEDTARTANLAEIDATEVEGRACAGCTI---SPGIVHYQFTA--GVPSGKSRSIQ 992
Db 788 VCGETWRRVAAVSSSIVYSDMKIRATRSVLKDEPHISSAKVLYDGVPGCGKTKELIS 847
QY 993 G---DVVVVVVPTRELNSMR--GFAAFTPHTAARV-----TIGR-----RVVID 1034
Db 848 RVNDEDEDVLYVPGQAAEMIRRRANSGLIATKENVRYTDSFLMNYGRGPCQYKRLFLD 907
QY 1035 EAPSLRPHLL--LLHMQRASSVHLLDPPNOIPAIID--FEHAGLVPAIRPELAPTSWMX 1088
Db 908 EGLMLHRCGVNPLVGMSLCSFAFYGDIOQIPIYINRAVTFPYPKHLSOLEVDAVETR--R 965
QY 1089 VTHRCPADVCELLRGAVP-KIOTTSRVLSLFW-----NEPAIQOKLVXTOAA 1135
Db 966 TTLRCPADITFEFLNOKYEGQWCTSVTRSVSHEVIOGAAMNPYSKPLKGVITFTQSD 1025
QY 1136 KA-----ANPQATVHEQOGATFTETITII-ATADARGIQQSSRAHAIVALTREITEK-- 1185
Db 1026 KSLLSRGYEDVHTVHEVOGETFEDVSLVRLTPPVGLISKQSPHLVLSLRHTRSIKYY 1085
QY 1186 CVILDA--PGLFREVG--ISDVIVN-----NPLAGGEVGHHRPSVI 1223
Db 1086 TVVIDAVVSVIRDLKCVSSYLLDMKVDVSTQOQIESVYKGVNLFVAAPKGTG----- 1138
QY 1224 PRGNPDQNLGTLQAFPPSCQ-----ISAVHQAELGHRPAPVAAVLPCCPELEOGLLY 1277
Db 1139 -----DVSDMOYYVDKCLPGNSTILNEYDAVTMOI-----1168
QY 1278 MPQELTVSDSVLVF-----ELNDIVHCRMAAPSQRK-AVLSTLYGRYGR--1321
Db 1169 RENSINWKCVDLMSKSVPLPRESETTLKPIRTAAEKPRKGLLENLVAMIKRNFNSPE 1228
QY 1322 -----TKYLEAHSADVRESLARFIPTIGPVQATTCCELYELVEAMWEEKGD 1366
Db 1229 LVGVVDIEDTASLVVDKFFD-AYLKEKKKRNIPILSRAS-----LERWIEK-QE 1277
QY 1367 GSAYLEL-----DLGNRDVSRITFEQKXCNKFTTGETIAHKVQOGISAMSKTECALFG 1420
Db 1278 KSTIGQLADFDIDLPAVDQYRHMIIKOOPKORLDLSIOTEYRAL-QTIVYHSKKINALFG 1336
QY 1421 PMFRAIEKEILLALPPN--IFYDAYEESV--FAAANVSAGSCMWFENDSEFDSTONNE 1476
Db 1337 PVSELTROLLETIDSSRFMYTRKTPQIIEFPSSDLSNVPMIDLELIDISKYDKSONEF 1396
QY 1477 SLGLECYVMECCGMPQWLIRLYHVRSAWILQAPKESILKGF-----WKXSGEGCT 1527
Db 1397 HCAVEYIWKRLGIDTFLAEY-----WKGHKRTLLKDYTAGIKTCLWYORKSGDVT 1449
QY 1528 LWNVTVMMAAIIACHYEFDFRVAFAFGDSDVVL--SDYRQSRNMAALIAAGCLKLKV 1584
Db 1450 FIGNTIIIACTLSMLPMERLIKAFGCGDSILYFPKGTDEPDIQOGANLL--WNFEAKL 1507

```

```

QY 1585 DYRPIGLYAGVVA--PGIGTLPDVVAFRAGRLSEKMWGSPGERAEOLRLAVCDFLRGLT 1641
Db 1508 FRKRYGFCGRYIIHHDRGCIYVYDPLKLSKLGAKHT-KNREHLEERTSTLCDVAGSLN 1566
QY 1642 NVAQV--CVDVYSRIVYGVSP 1659
Db 1567 NCAYYTHLNDAVGEVYKTPAP 1586

```

Search completed: May 30, 2001, 16:10:33
Job time: 215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:06:23 ; Search time 20.39 Seconds
(without alignments)
1599.806 Million cell updates/sec

Title: US-09-468-147-91

Sequence: 1 PGITTAIEQALAAANSALA.....FTETIKPVLDLNTSIQRVE 1698

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/pdata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/2/1aa/6A.COMB.pep:*
3: /cgn2_6/pdata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/pdata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/pdata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/pdata/2/1aa/6B.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7448	82.6	1693	4	US-08-478-507-7
2	7439	82.5	1693	3	US-08-840-316-1
3	7439	82.5	1693	4	US-08-809-523-1
4	7439	82.5	1693	5	PCT-US93-08849A-1
5	7439	82.5	1693	5	PCT-US93-08849-1
6	2057	22.8	431	4	US-08-478-507-2
7	515	5.7	1704	4	US-08-485-355B-40
8	374.5	4.2	2205	1	US-08-093-453B-2
9	365	4.0	78	2	US-07-876-941A-31
10	233	2.6	2500	2	US-08-801-263A-2
11	233	2.6	2500	3	US-09-102-248-2
12	218	2.4	2512	2	US-08-801-263A-9
13	218	2.4	2512	3	US-09-102-248-9
14	212.5	2.4	2517	2	US-08-801-263A-5
15	212.5	2.4	2517	3	US-09-102-248-5
16	211.5	2.3	2161	3	US-09-081-320-3
17	200	2.2	2431	1	US-07-920-281C-2
18	200	2.2	2431	4	US-08-466-277-2
19	184	2.0	1390	2	US-08-770-544-2
20	178	2.0	34	2	US-07-876-941A-32
21	165.5	1.8	255	3	US-09-081-320-14
22	162	1.8	1456	1	US-08-803-973-2
23	162	1.8	1456	1	US-08-803-972-2
24	152.5	1.7	459	4	US-09-080-983-5
25	148.5	1.6	3567	2	US-07-642-734C-4
26	148.5	1.6	3567	3	US-08-439-009A-3
27	145	1.6	1463	1	US-08-157-005-3

28	145	1.6	1463	4	US-08-747-863-3	Sequence 3, Appl
29	145	1.6	3724	2	US-08-804-227C-10	Sequence 10, Appl
30	145	1.6	3724	2	US-08-804-198-4	Sequence 4, Appl
31	143.5	1.6	1315	4	US-08-899-595-3	Sequence 3, Appl
32	140.5	1.6	1255	2	US-09-080-897-4	Sequence 4, Appl
33	140.5	1.6	1255	4	US-08-899-595-1	Sequence 1, Appl
34	140.5	1.6	1255	4	US-09-323-735-4	Sequence 4, Appl
35	140	1.6	2152	4	US-09-036-987A-3	Sequence 3, Appl
36	139	1.5	902	1	US-08-396-479B-6	Sequence 6, Appl
37	139	1.5	921	1	US-08-818-823-6	Sequence 2, Appl
38	136.5	1.5	921	1	US-08-396-479B-2	Sequence 2, Appl
39	136.5	1.5	921	1	US-08-818-823-2	Sequence 2, Appl
40	136	1.5	7257	4	US-09-335-409-5	Sequence 5, Appl
41	135.5	1.5	365	4	US-09-113-309-2	Sequence 2, Appl
42	134.5	1.5	4472	2	US-08-804-227C-2	Sequence 2, Appl
43	133.5	1.5	4302	3	US-08-658-136-5	Sequence 5, Appl
44	132	1.5	1248	2	US-09-080-897-2	Sequence 2, Appl
45	132	1.5	1248	4	US-09-323-735-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-478-507-7
Sequence 7, Application US/08478507
Patent No. 6120988
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbrough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988

```

ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-507-7

Query Match      82.6%  Score 7448:  DB 4;  Length 1693;
Best Local Similarity 81.8%:  Pred. No. 0;
Matches 1396;  Conservative 110;  Mismatches 169;  Indels 32;  Gaps 6;

QY 1  PGIITATEQALAAANSALANAAVVPFELSNVQTEIILINLMQPRQVLEFVLMNHP10R
    10 PGIITATEQALAAANSALANAAVVPFELSNVQTEIILINLMQPRQVLEFVLMNHP10R
    69
QY 61 VINHELEQYCARAGRCLENGAMPSTINDNPVLAHRCFLRPGVDORWTSA1PTGRPAAN
    70 VINHELEQYCARAGRCLENGAMPSTINDNPVLAHRCFLRPGVDORWTSA1PTGRPAAN
    129
QY 121 CRRSALGRLPADRTYCFDGSRCAPAEETGVALYSLHDLMPAVVAEMAHGXTRL1YAA
    130 CRRSALGRLPADRTYCFDGSRCAPAEETGVALYSLHDLMPAVVAEMAHGXTRL1YAA
    189
QY 181 LHLPEVLLPRTYHTTSYLLIHNDGRAVVTYEGDTSAGYNHDSYILRAMIRTKIYGDH
    190 LHLPEVLLPRTYHTTSYLLIHNDGRAVVTYEGDTSAGYNHDSYILRAMIRTKIYGDH
    249
QY 241 PLVIERRAICGHVLLLTAPAPSPMPYVPRSTEVYVNSITGPGSGSELEFSSACSTK
    250 PLVIERRAICGHVLLLTAPAPSPMPYVPRSTEVYVNSITGPGSGSELEFSSACSTK
    309
QY 301 STFAVAVNHIMDRMLFGATLDQAFCCSRMLTYLRGISYVTVGVALYANGGMASEDAL
    310 STFAVAVNHIMDRMLFGATLDQAFCCSRMLTYLRGISYVTVGVALYANGGMASEDAL
    369
QY 361 TAYITAAVLTICHOVYLRTOAISKGMRLGVEHAKFTIRLYSWLFEKSGRDY1PGROLO
    370 TAYITAAVLTICHOVYLRTOAISKGMRLGVEHAKFTIRLYSWLFEKSGRDY1PGROLO
    429
QY 421 FYACCRWMLSGFHLDPRLVYFEDSVPCRCRTFLKAYAGKCCCMRMNLGOCCTGLEPAE
    430 FYACCRWMLSGFHLDPRLVYFEDSVPCRCRTFLKAYAGKCCCMRMNLGOCCTGLEPAE
    489
QY 481 GLYGDHNDNEAYEGSEVDEPAEPALHDVSGTYAVAHGQLEALYRALVNPQDIAARSLRT
    490 GAVGDGHDNDNEAYEGSEVDEPAEPALHDVSGTYAVAHGQLEALYRALVNPQDIAARSLRT
    549
QY 541 ATYVELVASPDLRECRYLGNKTFPTTVVQGAHNEANGPEQVYLSFDAOSMGAGSHSLT
    550 ATYVELVASPDLRECRYLGNKTFPTTVVQGAHNEANGPEQVYLSFDAOSMGAGSHSLT
    609
QY 601 YELTPAGLOVRISSNGLDCTATPPPGAGSAPAEVAAFCFALYRYARFQORHSLTGGLW
    610 YELTPAGLOVRISSNGLDCTATPPPGAGSAPAEVAAFCFALYRYARFQORHSLTGGLW
    669
QY 661 LHEBGLIGIPPPSPGHIMESANPFCGEGTLTTRTWS-TSGFSSDSESP-----PEAAPA
    670 FHEBGLIGIPPPSPGHIMESANPFCGEGTLTTRTWS-TSGFSSDSESP-----PEAAPA
    729
QY 715 MAATPGILPHSTPPVSDIWLVPPESEFQVDAAYV-PPAPDPAGIL-GRVVLTPPPPPVH
    730 MAATPGILPHSTPPVSDIWLVPPESEFQVDAAYV-PPAPDPAGIL-GRVVLTPPPPPVH
    773
QY 773 KPSIP-PPSNNRRLTYTPPGAKYVAGSLFESDCDMLVMSNPGHRGGGGLCHAFYORFP
    831

```

```

Db 767 APAITHOTARRHRLFTYIPDGSKYFAGSLSESTCTWLVNMSNVDBRGCGGLCHAFYORFP
QY 832 EAFYPTFEIMEGLAATYLPRLPIIHAVAPDYREQONPKRLEAYRRTCSRGTAAYPL
    833 EAFYPTFEIMEGLAATYLPRLPIIHAVAPDYREQONPKRLEAYRRTCSRGTAAYPL
    891
Db 827 ASFDAASFVMDGAATYLPRLPIIHAVAPDYREQONPKRLEAYRRTCSRGTAAYPL
    828 ASFDAASFVMDGAATYLPRLPIIHAVAPDYREQONPKRLEAYRRTCSRGTAAYPL
    886
QY 892 GSGIYQVPSLSPDAMERNRHPGDELTLTEPAWMPFANKPAQVLTITTEDTARTANLAL
    893 GSGIYQVPSLSPDAMERNRHPGDELTLTEPAWMPFANKPAQVLTITTEDTARTANLAL
    951
Db 887 GTGIYQVPIGSPFAMERNRHPGDELTLTEPAWMPFANKPAQVLTITTEDTARTANLAL
    888 GTGIYQVPIGSPFAMERNRHPGDELTLTEPAWMPFANKPAQVLTITTEDTARTANLAL
    946
QY 952 EIDATEVGRACAGCTTSPGIVHIOFTAGVPGSGKSSIOGGDVVVVPTRELNSMRR
    953 EIDATEVGRACAGCTTSPGIVHIOFTAGVPGSGKSSIOGGDVVVVPTRELNSMRR
    1011
Db 947 ELDSATDVGRACAGCTTSPGIVHIOFTAGVPGSGKSSIOGGDVVVVPTRELNSMRR
    948 ELDSATDVGRACAGCTTSPGIVHIOFTAGVPGSGKSSIOGGDVVVVPTRELNSMRR
    1006
QY 1012 RGFAPFTPHTAARVLTIGRRVYIDEAPSLPHLLLLHMOARASVHLGDDPNQIPIADEHA
    1013 RGFAPFTPHTAARVLTIGRRVYIDEAPSLPHLLLLHMOARASVHLGDDPNQIPIADEHA
    1071
Db 1007 RGFAPFTPHTAARVLTIGRRVYIDEAPSLPHLLLLHMOARASVHLGDDPNQIPIADEHA
    1008 RGFAPFTPHTAARVLTIGRRVYIDEAPSLPHLLLLHMOARASVHLGDDPNQIPIADEHA
    1066
QY 1072 GLVPAIRPELAPTSWNVTRHRCPADVCELIRGAYPKIOTTSRYVLSLFWNEPAIGQRLVX
    1073 GLVPAIRPELAPTSWNVTRHRCPADVCELIRGAYPKIOTTSRYVLSLFWNEPAIGQRLVX
    1131
Db 1067 GLVPAIRPELAPTSWNVTRHRCPADVCELIRGAYPKIOTTSRYVLSLFWNEPAIGQRLVX
    1068 GLVPAIRPELAPTSWNVTRHRCPADVCELIRGAYPKIOTTSRYVLSLFWNEPAIGQRLVX
    1126
QY 1132 TOAKKANPGCAITVHEAGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA
    1133 TOAKKANPGCAITVHEAGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA
    1191
Db 1127 TOAKKANPGCAITVHEAGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA
    1128 TOAKKANPGCAITVHEAGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA
    1186
QY 1192 PGLLRREVGISDVYVNNFNLGGEVGHXRPSTYIPGPNPDONIGTQAPPPSCQISAYHQLA
    1193 PGLLRREVGISDVYVNNFNLGGEVGHXRPSTYIPGPNPDONIGTQAPPPSCQISAYHQLA
    1251
Db 1187 PGLLRREVGISDVYVNNFNLGGEVGHXRPSTYIPGPNPDONIGTQAPPPSCQISAYHQLA
    1188 PGLLRREVGISDVYVNNFNLGGEVGHXRPSTYIPGPNPDONIGTQAPPPSCQISAYHQLA
    1246
QY 1252 EELGHRPAPVAAVLPCEPELEQGLYMPQELTVSDVSLVELTDIVHCRMAAPSQRKAVL
    1253 EELGHRPAPVAAVLPCEPELEQGLYMPQELTVSDVSLVELTDIVHCRMAAPSQRKAVL
    1311
Db 1247 EELGHRPAPVAAVLPCEPELEQGLYMPQELTVSDVSLVELTDIVHCRMAAPSQRKAVL
    1248 EELGHRPAPVAAVLPCEPELEQGLYMPQELTVSDVSLVELTDIVHCRMAAPSQRKAVL
    1306
QY 1312 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTEGELVELYAMNEKQDGSAYL
    1313 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTEGELVELYAMNEKQDGSAYL
    1371
Db 1307 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTEGELVELYAMNEKQDGSAYL
    1308 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTEGELVELYAMNEKQDGSAYL
    1366
QY 1372 ELDLCNNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKEIL
    1373 ELDLCNNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKEIL
    1431
Db 1367 ELDLCNNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKEIL
    1368 ELDLCNNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKEIL
    1426
QY 1432 ALPPIIFGYDAEESVFAAASGAGSCWVPEENDESEFSDSTONNFSGLCEVYNEECGMP
    1433 ALPPIIFGYDAEESVFAAASGAGSCWVPEENDESEFSDSTONNFSGLCEVYNEECGMP
    1491
Db 1427 ALPPIIFGYDAEESVFAAASGAGSCWVPEENDESEFSDSTONNFSGLCEVYNEECGMP
    1428 ALPPIIFGYDAEESVFAAASGAGSCWVPEENDESEFSDSTONNFSGLCEVYNEECGMP
    1486
QY 1492 QMLIRLYHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDPRVA
    1493 QMLIRLYHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDPRVA
    1551
Db 1487 QMLIRLYHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDPRVA
    1488 QMLIRLYHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDPRVA
    1546
QY 1552 AFKGDSDVLSVCRQSRNMAALJAGCGLKLYDRPIGTYAGVVAVPAGLTPDVRFA
    1553 AFKGDSDVLSVCRQSRNMAALJAGCGLKLYDRPIGTYAGVVAVPAGLTPDVRFA
    1611
Db 1547 AFKGDSDVLSVCRQSRNMAALJAGCGLKLYDRPIGTYAGVVAVPAGLTPDVRFA
    1548 AFKGDSDVLSVCRQSRNMAALJAGCGLKLYDRPIGTYAGVVAVPAGLTPDVRFA
    1606
QY 1612 GRUSEKNWGGPERABDLRLAVCDLRLGTLNNAOVCVVVSRYVGSBGLVHNLIGMLQOT
    1613 GRUSEKNWGGPERABDLRLAVCDLRLGTLNNAOVCVVVSRYVGSBGLVHNLIGMLQOT
    1671
Db 1607 GRILEKNWGGPERABDLRLAVCDLRLGTLNNAOVCVVVSRYVGSBGLVHNLIGMLQOT
    1608 GRILEKNWGGPERABDLRLAVCDLRLGTLNNAOVCVVVSRYVGSBGLVHNLIGMLQOT
    1666
QY 1672 IADGKAHETETIKPVLDLTNSIIORVE
    1673 IADGKAHETETIKPVLDLTNSIIORVE
    1698
Db 1667 IADGKAHETETIKPVLDLTNSIIORVE
    1668 IADGKAHETETIKPVLDLTNSIIORVE
    1693

RESULT 2
US-08-840-316-1
Sequence 1, Application US/08840316
Patent No. 6034567
GENERAL INFORMATION:
APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
APPLICANT: Tsarev, Sergei. A., and Robinson, Robin A.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their

```

```

: TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
: NUMBER OF SEQUENCES: 111
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/840.316
: FILING DATE: 11-APR-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Richard W. Bork
: REGISTRATION NUMBER: 36,459
: REFERENCE/DOCKET NUMBER: 2026-4255
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 751-4800
: TELEFAX: (212) 751-6849
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1693 AMINO ACID RESIDUES
: TYPE: AMINO ACID
: STRANDEDNESS: UNKNOWN
: TOPOLOGY: UNKNOWN
: US-08-840-316-1

```

Query Match 82.5%; Score 7439; DB 3: Length 1693;
 Best Local Similarity 81.6%; Pred. No. 0;
 Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

```

QY 1 PGTTAIEBOAALAAANSALANAVYVPRPLSRVOTELLINIMOPROLVFREVLMNPIOR 60
DB 10 PGTTAIEBOAALAAANSALANAVYVPRPLSHQIETILINMOPROLVFREVLMNPIOR 69
QY 61 VIHNELEYOCRAVAGRCLEVGARHSINDPNVLRHCFLPVGRDQVQWTSAPTRGPAN 120
DB 70 VIHNELEYOCRAVAGRCLEVGARHSINDPNVLRHCFLPVGRDQVQWTSAPTRGPAN 129
QY 121 CRRSALRGLPADRYCFCDFGSCAFPAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
DB 130 CRRSALRGLPADRYCFCDFGSCAFPAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 189
QY 181 LHPPEVLLPFGYVHTSYLLIHDSGRAYVYTGDTSGAGNHVSLIRAMIRTKTYGDH 240
DB 190 LHPPEVLLPFGYVHTSYLLIHDSGRAYVYTGDTSGAGNHVSLIRAMIRTKTYGDH 249
QY 241 PLVIEVERAIGCHFVLLTAAPESPMPYVPRSTEVYRSIFGSGSPSLPSPACSTK 300
DB 250 PLVIEVERAIGCHFVLLTAAPESPMPYVPRSTEVYRSIFGSGSPSLPSPACSTK 309
QY 301 STFHANVPIHMDLRFGLTLDQAFCCSRMLMTYLGISTKYTVGLVLANEGNASEDAL 360
DB 310 STFHANVPIHMDLRFGLTLDQAFCCSRMLMTYLGISTKYTVGLVLANEGNASEDAL 369
QY 361 TAITTAAYLTICHOYRLTQOAIKGMRLGVEHAOKFITRLYSMLPEKSGRDVTPGRLO 420
DB 370 TAITTAAYLTICHOYRLTQOAIKGMRLGVEHAOKFITRLYSMLPEKSGRDVTPGRLO 429
QY 421 FYAQCRRLISAGEHLDPRVLVPEDESVCRCRTEFLKRVAGFCFCFMRMLGOECTFLEPAE 480
DB 430 FYAQCRRLISAGEHLDPRVLVPEDESAPCHCRTRAIRKAVSFFCCFMMMLGECCTCFIQPAE 489

```

```

QY 481 GLYGDGHNEAYESSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPQDITAAASRLT 540
DB 490 GYVGDGHNEAYESSEVDPAEPASIDISGSYVPGTALQPLQALDLPETIYARGRILT 549
QY 541 ATVEIASPDRLCECTVLGKTEFTTVVDGAHLEANGPEQYVLSFASRQSMGASHSLT 600
DB 550 ATYKVSQVGRIDCELTLLKNTFTTSFVDGAVLETNGPERHNSLFPASOSTMAAGFSILT 609
QY 601 YELTPAGLOYRISSNGLDCTATFPFGAPASAPAEVAALFCSALYRYNRTQRHSLTGILM 660
DB 610 YASAAGLEVRYVAAGLDHRAVAPAGVSPRASAGEVTAFCASALYRNRBAQRLSLTGIMF 669
QY 661 LHPGGLGIFPPSPGHINESANPFCGEGILYRTMS-----TSGFSDSPPE 709
DB 670 FHPGGLGIFPPSPGHINESANPFCGEGILYRTMS-----TSGFSDSPPE 724
QY 710 AAPAPAAATPPGLPHSTPPVSDIWLPPPSDEFQVDAAPV-PPAPDBA-GLPGVVLTPPP 767
DB 725 PSIPSRAATP-----TP-----AAPLPAPADPSPTLSAPARGBAP 761
QY 768 PPEVHRPSIP-PPSRNRRLTYTPDGAKYAGSLFESDCDWLVNANPGHRRPGGLCHAF 826
DB 762 GATAPAPAITHTQARHRLITFYPDGSKYFAGSLFESTCTWLVNANVDRHPGGGLCHAF 821
QY 827 YORPPAFTPTERTREGLAATYLTTPRPIIHAVAPDYRVQONPKRLDAAYRETCSRGA 886
DB 822 YORPPAFTPTERTREGLAATYLTTPRPIIHAVAPDYRVQONPKRLDAAYRETCSRGA 881
QY 887 AYPGLSGIYQVPSLSPDAMERNHRPGDELTYTEPAANFEANRKAOPVLTTETETAR 946
DB 882 AYPGLSGIYQVPSLSPDAMERNHRPGDELTYTEPAANFEANRKAOPVLTTETETAR 941
QY 947 ANLAIEIDATEVGRACAGCTISPGIVHYQFTAGVBSGSKRSIQGDVYVVVPTREL 1006
DB 942 ANLAIEIDATEVGRACAGCTISPGIVHYQFTAGVBSGSKRSIQGDVYVVVPTREL 1001
QY 1007 NSWRRRGPAFTPTHTARTYIGRRVYIDEAPSLPHILLHMRASSVHLGDPNQPAL 1066
DB 1002 NAMRRRGPAFTPTHTARTYIGRRVYIDEAPSLPHILLHMRASSVHLGDPNQPAL 1061
QY 1067 DFPHAGLVPAIRPDLAPTSWXYVTHRCPADVCYLIGAVPKIDTGRVLSLWGPENAG 1126
DB 1062 DFPHAGLVPAIRPDLAPTSWXYVTHRCPADVCYLIGAVPKIDTGRVLSLWGPENAG 1121
QY 1127 QKLVYTOAKKAANPGAITVHEAOGATFETTTIATADARGLIOSSRAHAVALTRHTEKC 1186
DB 1122 QKLVYTOAKKAANPGAITVHEAOGATFETTTIATADARGLIOSSRAHAVALTRHTEKC 1181
QY 1187 VIIDAPGLREVGISDIYNNFVLAGEVGHRRPSVYPRGNPDQNTGLQAFPPSCQISA 1246
DB 1182 VIIDAPGLREVGISDIYNNFVLAGEVGHRRPSVYPRGNPDQNTGLQAFPPSCQISA 1241
QY 1247 YHOLAEELGHRPAVAAVLPCEPELBOGLYMOQELTVPSVYFELTDIVHRMAAPSO 1306
DB 1242 YHOLAEELGHRPAVAAVLPCEPELBOGLYMOQELTVPSVYFELTDIVHRMAAPSO 1301
QY 1307 RKAVLSTLVGRYGRRTKLYEAHSDVRESLARPITPIGVOATTCELVLEAMAVEKGOD 1366
DB 1302 RKAVLSTLVGRYGRRTKLYEAHSDVRESLARPITPIGVOATTCELVLEAMAVEKGOD 1361
QY 1367 GSAVLELDLCNRDYSRTFFQKCKNFTTGETTAHKGVGOGISAWSKTFCALGPMFRAI 1426
DB 1362 GSAVLELDLCNRDYSRTFFQKCKNFTTGETTAHKGVGOGISAWSKTFCALGPMFRAI 1421
QY 1427 EKEILLALPPIINIFYGAYEESVFAAASGASCMVEENDESEDSQONNFSLEGVME 1486
DB 1422 EKEILLALPPIINIFYGAYEESVFAAASGASCMVEENDESEDSQONNFSLEGVME 1481
QY 1487 ECGMPQWLIRLYHLVRSAMILLQPKESLGFWRKSGSEPQTLLMNTVMNNAIIAHCEYFR 1546
DB 1482 ECGMPQWLIRLYHLVRSAMILLQPKESLGFWRKSGSEPQTLLMNTVMNNAIIAHCEYFR 1541

```



```

Db 1122 OKLVYTOAKANAPGVSIVHEAGATYETETIIIAFADARGLIOSSRAHAIVALTFRTEKC 1181
Qy 1187 VILDPGLREVGISDVIVNNEFLAGEVGHXHSVIRPGNDONLGLQAPPPSCQISA 1246
Db 1182 VILDPGLREVGISDAIVNNEFLAGEVGHXHSVIRPGNDONLGLQAPPPSCQISA 1241
Qy 1247 YHOLAEELGHRPAPAAVLPCCPELEOGLLYMPOELVSDSVIVELTDIYHCRMAAASQ 1306
Db 1242 FHEILAEELGHRPAPAAVLPCCPELEOGLLYMPOELVSDSVIVELTDIYHCRMAAASQ 1301
Qy 1307 RKAVALSTLVGRYGRRTKLYEAHSDVRESLARFIPTIGPVQATTCELVELEAWEKQOD 1366
Db 1302 RKAVALSTLVGRYGRRTKLYEAHSDVRESLARFIPTIGPVQATTCELVELEAWEKQOD 1361
Qy 1367 GSAVLELDLCNRDVSRTFFQKXCNKFTTGETIAHGKVGOGISAMSKTFICALFGWFRAT 1426
Db 1362 GSAVLELDLCNRDVSRTFFQKXCNKFTTGETIAHGKVGOGISAMSKTFICALFGWFRAT 1421
Qy 1427 EKELIALLPNIFFYGDAAEESVFAAASGAGSCWFEENDESEPDSTONNFSIGLECYME 1486
Db 1422 EKELIALLPNIFFYGDAAEESVFAAASGAGSCWFEENDESEPDSTONNFSIGLECYME 1481
Qy 1487 ECGMPQWIRLYHYRSAMIIQAPKESLKGFWKHSGEPGLTLMNTVMMAIIAHCEYER 1546
Db 1482 ECGMPQWIRLYHYRSAMIIQAPKESLKGFWKHSGEPGLTLMNTVMMAIIAHCEYER 1541
Qy 1547 DFRVAAPGDDSVLCSTYROSRAAIIAGGCLKYDPRIGLYAGVVAAPGIGLTPD 1606
Db 1542 DFRVAAPGDDSVLCSTYROSRAAIIAGGCLKYDPRIGLYAGVVAAPGIGLTPD 1601
Qy 1607 VYRFAGRLSEKMMGPGPRAEQLRLACDFLGLTFNVAQVCDDVYSRYGVSPGLVHNI 1666
Db 1602 VYRFAGRLSEKMMGPGPRAEQLRLACDFLGLTFNVAQVCDDVYSRYGVSPGLVHNI 1661
Qy 1667 GMLQTIADGKAHFTETIKPVLDTNLSIIQRYE 1698
Db 1662 GMLQTIADGKAHFTETIKPVLDTNLSIIQRYE 1693

```

RESULT 4
PCT-US93-08849A-1
Sequence 1, Application PC/TUS9308849A
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
OPERATING SYSTEM: IBM PC COMPATIBLE
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08849A
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US07/947,263
FILING DATE: 18-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: William S. Feller
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4032 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 AMINO ACID RESIDUES
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
PCT-US93-08849A-1

Query Match      82.5%; Score 7439; DB 5; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

Qy 1 PGITTAIQAALAAANSALANAVYVPRFLSKYQVEILNLQPROLVPRVYLANHPFOR 60
Db 10 PGITTAIQAALAAANSALANAVYVPRFLSKYQVEILNLQPROLVPRVYLANHPFOR 69
Qy 61 VIHNELEQYCARARAGCLVEGVAHPSINDNPNVLHRCFLRPVGRDVRWYSAPTRGPAN 120
Db 70 VIHNELEQYCARARAGCLVEGVAHPSINDNPNVLHRCFLRPVGRDVRWYSAPTRGPAN 129
Qy 121 CRRSALRGLPPADRTYCDGFSRCAFAETGVALYSLHDLMPADVAEAMARHGTRLYAA 180
Db 130 CRRSALRGLPPADRTYCDGFSRCAFAETGVALYSLHDLMPADVAEAMARHGTRLYAA 189
Qy 181 LHLPEVLLPQTYHTTLYLHGDRAVYVEGDTSGAGYHNDVSLAMITRTIYVDH 240
Db 190 LHLPEVLLPQTYHTTLYLHGDRAVYVEGDTSGAGYHNDVSLAMITRTIYVDH 249
Qy 241 PLVIERVRAIGCHFYLLTJAPEPSPMPYVPRSTEVYVRSIFGPGSPSLFPGSCTK 300
Db 250 PLVIERVRAIGCHFYLLTJAPEPSPMPYVPRSTEVYVRSIFGPGSPSLFPGSCTK 309
Qy 301 STEFAVPHVIMDRMLFCAVLDQAFCCSRMTYLRGISYKVTGALVANEGMNASDAL 360
Db 310 STEFAVPHVIMDRMLFCAVLDQAFCCSRMTYLRGISYKVTGALVANEGMNASDAL 369
Qy 361 TAXITAAVLTICHOXYLTQOATISKGMRLGVHACKFTIRLYSWLFEEGSGRYTGRLO 420
Db 370 TAXITAAVLTICHOXYLTQOATISKGMRLGVHACKFTIRLYSWLFEEGSGRYTGRLO 429
Qy 421 FYAOCRRMLSGAFHLDPRVLVDESVPCRCRTFLKVGAKGCCFMRMLGOECTCLEPAE 480
Db 430 FYAOCRRMLSGAFHLDPRVLVDESVPCRCRTFLKVGAKGCCFMRMLGOECTCLEPAE 489
Qy 481 GLVGDHGDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDITARASRLT 540
Db 490 GLVGDHGDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDITARASRLT 549
Qy 541 ATVELVAPDRLEQTVLGKNTFRITVVDGAILENGPEQVYLSRDSAROSGASHSLT 600
Db 550 ATVELVAPDRLEQTVLGKNTFRITVVDGAILENGPEQVYLSRDSAROSGASHSLT 609
Qy 601 YELTPAGLOVRITSSNGLDCTATFPPGAPASAPAGEVAFCALYRNFRTORHSLTGLM 660
Db 610 YELTPAGLOVRITSSNGLDCTATFPPGAPASAPAGEVAFCALYRNFRTORHSLTGLM 669
Qy 661 LAPEGLLIFPPSPGHIMESANPFCGEGTLYTRWS-----TSGFSSDSFSPE 709
Db 670 LAPEGLLIFPPSPGHIMESANPFCGEGTLYTRWS-----TSGFSSDSFSPE 724
Qy 710 AAPAPMAATPGIPLHSTPPYSDIWTVLPPSEEFQVYNAAYV-PAAPPA-GLQPPVLTLP 767
Db 725 PPSIPRAATP-----TP-----AAPPPAPPSPLSLAPARGEP 761
Qy 768 PPPVHKPSIP-PPSRNRRLTYTPDGAKYAGSLFESDCDMLVNSNPGHRGGGLCAH 826
Db 762 GATAPAPATHTOTAHRRLLFTYPPGSKYFAGSLFESTCTWLYNNSNDHRGGGLCAH 821
Qy 827 YORPEEAFYPTFEIMREGILAAVTLPRPIIHAVADRYEONPKRLEAAYRETCSRGT 886
Db 822 YORPEEAFYPTFEIMREGILAAVTLPRPIIHAVADRYEONPKRLEAAYRETCSRGT 881

```

```

QY 887 ATPELLSGITQVSVLSFDMERNHPRGDELTYLTERPANNFEANKPAQVLTITTEDTART 946
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 882 ATPLLGITQVSVLSFDMERNHPRGDELTYLTERPANNFEANKPAQVLTITTEDTART 941
QY 947 ANALEIDATEVGRACAGCTISGIVHOFTAGVPGSGSRISIOGDDVVVVVPPRELR 1006
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 942 ANALEIDATEVGRACAGCTISGIVHOFTAGVPGSGSRISIOGDDVVVVVPPRELR 1001
QY 1007 NSMRRGFAFTPHRTARVIGRRVVIDEAPSLPHILLHLMORASSVHLGDPNQIPAI 1066
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1002 NAMRRRGFAFTPHRTARVIGRRVVIDEAPSLPHILLHLMORASSVHLGDPNQIPAI 1061
QY 1067 DEFHAGLVPAIRBELAFTSWXVYTHRCPADVCELRIGATYKIOYTSVYLSLFWNEPAIG 1126
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1062 DEFHAGLVPAIRBELAFTSWXVYTHRCPADVCELRIGATYKIOYTSVYLSLFWNEPAIG 1121
QY 1127 OKLVYTOAANAAMPATVIEAGAFTEETTTATADARGLIOSRRAHALVALTRHEKC 1186
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1122 OKLVYTOAANAAMPATVIEAGAFTEETTTATADARGLIOSRRAHALVALTRHEKC 1181
QY 1187 VILDAEGLREVGISDVIVNFFLAGEVGXRHPSPVIRGNPNQNLGTOAFPPSCQISA 1246
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1182 VILDAEGLREVGISDVIVNFFLAGEVGXRHPSPVIRGNPNQNLGTOAFPPSCQISA 1241
QY 1247 YHOLAEELGRAPVPAVAVLPCEPELEGILLYMPQELTVSDSVLVFELTDIVHCRMAAPSQ 1306
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1242 YHOLAEELGRAPVPAVAVLPCEPELEGILLYMPQELTVSDSVLVFELTDIVHCRMAAPSQ 1301
QY 1307 RKAVLSTIVRGVRRTKLYAASHDVRESLARPTIGPQATTCCLXELVEAMVERGQD 1366
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1302 RKAVLSTIVRGVRRTKLYAASHDVRESLARPTIGPQATTCCLXELVEAMVERGQD 1361
QY 1367 GSAVLELDLCNRDVSRTTFPOKXCNKFTTGETTIAHCKVGGISAMSKTFCALGPMFRAI 1426
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1362 GSAVLELDLCNRDVSRTTFPOKXCNKFTTGETTIAHCKVGGISAMSKTFCALGPMFRAI 1421
QY 1427 EKEIALLPBNIFYDAYEESVFAAASGASCMVFENDESEFDSQONNESLGLCYVME 1486
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1422 EKEIALLPBNIFYDAYEESVFAAASGASCMVFENDESEFDSQONNESLGLCYVME 1481
QY 1487 ECGMPOMLRLYLHVSAMTLOAPKESLKGFMKHSGEPOTLLMNTVWNAITIAHCYEF 1546
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1482 ECGMPOMLRLYLHVSAMTLOAPKESLKGFMKHSGEPOTLLMNTVWNAITIAHCYEF 1541
QY 1547 DFRVAFAKGGDSVVLCSYDYSRNSNAALJAGCGLKLVDRPIGLYAGVVAAGLGLTLPD 1606
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1542 DFRVAFAKGGDSVVLCSYDYSRNSNAALJAGCGLKLVDRPIGLYAGVVAAGLGLTLPD 1601
QY 1607 VYFAFARLSKMMGPRPERAEQRLAVCPDLRGITNAOYCVVYVSRVYSGVLVHNL 1666
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1602 VYFAFARLSKMMGPRPERAEQRLAVCPDLRGITNAOYCVVYVSRVYSGVLVHNL 1661
QY 1667 GMLQTIADGKAHFTETIKPYLDLTNSIOHVE 1698
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1662 GMLQTIADGKAHFTETIKPYLDLTNSIOHVE 1693

```

RESULT 5

PCT-US93-08849-1
Sequence 1, Application PC/TUS9308849

GENERAL INFORMATION:

APPLICANT: Tsarev, Sergei A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA

```

      ? ZIP: 10154
      ? COMPUTER READABLE FORM:
      ? MEDIUM TYPE: FLOPPY DISK
      ? OPERATING SYSTEM: IBM PC COMPATIBLE
      ? SOFTWARE: WORDPERFECT 5.1
      ? CURRENT APPLICATION DATA:
      ? FILING DATE: 17-SEP-1993
      ? CLASSIFICATION:
      ? PRIOR APPLICATION DATA:
      ? APPLICATION NUMBER: 07/947,263
      ? FILING DATE: 18-SEP-1992
      ? NAME:
      ? ATTORNEY/AGENT INFORMATION:
      ? NAME: Bork, Richard, W.
      ? REGISTRATION NUMBER: 36,459
      ? REFERENCE/DOCKET NUMBER: 2026-4032
      ? TELECOMMUNICATION INFORMATION:
      ? TELEPHONE: (212) 758-4800
      ? TELEFAX: (212) 751-6849
      ? INFORMATION FOR SEQ. ID NO. 1:
      ? SEQUENCE CHARACTERISTICS:
      ? LENGTH: 1693 amino acid residues
      ? TYPE: amino acid
      ? STRANDEDNESS: unknown
      ? TOPOLOGY: unknown
      ? PCT-US93-08849-1

```

Query Match 82.5%; Score 7439; DB 5; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

```

QY 1 PGTTTAEQAAALAAASALANAVVPRPFLSRVGTETLLINMOPROLYVPRPEVLMNPIOR 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 PGTTTAEQAAALAAASALANAVVPRPFLSHQIEILLINMOPROLYVPRPEVLMNPIOR 69
QY 61 VINHELEVCRAAGSCLEVGAPRISINDPNVLRHCFLPVGRDVOQRYMSAPTRGPAAN 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 70 VINHELEVCRAAGSCLEVGAPRISINDPNVLRHCFLPVGRDVOQRYMSAPTRGPAAN 129
QY 121 CRRSALRGLEPPADRYCFCDFGFSKCAFAETGVALYSLHDLMPADVAEAMARHCXRLVYA 180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 CRRSALRGLEPPADRYCFCDFGFSKCAFAETGVALYSLHDLMPADVAEAMARHCXRLVYA 189
QY 181 LHLPEVLLPBGVYHTSVLLIHDGRVAVTYEGDTSAGINHDSVSTLRAMIRTKTVGGH 240
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 LHLPEVLLPBGVYHTSVLLIHDGRVAVTYEGDTSAGINHDSVSTLRAMIRTKTVGGH 249
QY 241 PLVIERVRAIGCHFVLLTAADEPSPMPVYPVPRSTEVYVRSIFGSGSPSLFPASCTK 300
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 PLVIERVRAIGCHFVLLTAADEPSPMPVYPVPRSTEVYVRSIFGSGSPSLFPASCTK 309
QY 301 STFHAVPIHIMRLMFGATLDDQAFCCSRMLMTYLRGISKYTVGALVYANEGMNASEDL 360
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 STFHAVPIHIMRLMFGATLDDQAFCCSRMLMTYLRGISKYTVGALVYANEGMNASEDL 369
QY 361 TAITTAAYLTICHOURLFOAISKGRRLGVEHAOKFETIRLTSWLFKSGGRYIIPROLO 420
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 370 TAITTAAYLTICHOURLFOAISKGRRLGVEHAOKFETIRLTSWLFKSGGRYIIPROLO 429
QY 421 FYAOCRRMLASGFHLDPRVLVFEDESVCRCRTFLKVKVAKFCFCFMMWLOEOTCELEPAE 480
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 430 FYAOCRRMLASGFHLDPRVLVFEDESVCRCRTFLKVKVAKFCFCFMMWLOEOTCELEPAE 489
QY 481 GLVGDHNDNEAYEGSEVDPAEPAHLDSGYTYAHGHOLEALYRALNVPQDIYAARSLT 540
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 GLVGDHNDNEAYEGSEVDPAEPAHLDSGYTYAHGHOLEALYRALNVPQDIYAARSLT 549
QY 541 ATVEVLAAPDLRCRTVGNKTFRTTVVDGAHLNENGPQOYVLSFPAQSQSGAGSHSLT 600
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 550 ATVEVLAAPDLRCRTVGNKTFRTTVVDGAHLNENGPQOYVLSFPAQSQSGAGSHSLT 609

```


Matches	381: Conservative	26: Mismatches	24: Indels	0: Gaps
QY	1257	RAPAAVAVLPPEPELEQGLLMPQELTVSDSVLPELTDIVICRMAAPQQRKAVLSTLVG	1316	
Db	1	RVPVAAPVLPPEPELEQGLLMPQELTVSDSVLPELTDIVICRMAAPQQRKAVLSTLVG	60	
QY	1317	RGRGRTKLYEAASHDVRESLARFIPTIGVQATCELYELVAMVEKGQDGSAVLELDLC	1376	
Db	61	RGGGRTKLYNASHSDVRESLARFIPTIGVQATCELYELVAMVEKGQDGSAVLELDLC	120	
QY	1377	NBDVSRITFFQKXCKNFETTGETIAGHKVQGGISAMSKFTCALFGWFRATIEKILLALPP	1436	
Db	121	NBDVSRITFFQKXCKNFETTGETIAGHKVQGGISAMSKFTCALFGWFRATIEKILLALPP	180	
QY	1437	NIFYDAYEESYFAAASVAGSCGCMYFENFSEFSDTONNFSGLGECVWVEECCMPOMLR	1496	
Db	181	GVFYGDADPDYTFSAVAAVAAKASMYFENFSEFSDTONNFSGLGECVWVEECCMPOMLR	240	
QY	1497	LYHLVRSAMIIQAPKESLKGFWKHSGBEGTLLMNTVMNAIIAHCEYRDFRVAAFKGD	1556	
Db	241	LYHLRSAMIIQAPKESLKGFWKHSGBEGTLLMNTVMNAIYTHCYDPRQVAAFKGD	300	
QY	1557	DSVVLCSVDROSRMAALLAGCGLKLVXYRRTIGYAGVVAAPGLTLPDVRVFRGRUSE	1616	
Db	301	DSVVLCSVRSQSPGAVALLAGCGLKLVDFRPIGLYAGVVAAPGLTLPDVRVFRGRUSE	360	
QY	1617	KNMGGPERAEQDLRLAVCDFELGTNNVAVQVADVVSRYVSGPGLVHNLGMLQTIADSK	1676	
Db	361	KNMGGPERAEQDLRLAVSDFLRKLRTNVAGCMQVDVVSRYVSGPGLVHNLGMLQVADSK	420	
QY	1677	AHFTETIKRVL 1687		
Db	421	AHFTESVAKRVL 431		
<p>RESULT 7 US-08-485-355B-40 Sequence 40 Application US/08485355B Patent No. 6177075</p> <p>GENERAL INFORMATION: APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N. TITLE OF INVENTION: Insect Viruses and their Uses in Protecting Plants</p> <p>NUMBER OF SEQUENCES: 57 CORRESPONDENCE ADDRESS: ADDRESSEE: Flehr Hombach Test Albritton & Herbert LLP STREET: Four Embardadero Center, Suite 3400 CITY: San Francisco STATE: California COUNTRY: United States ZIP: 94111-4187</p> <p>COMPUTER READABLE FORM: MEDIUM TYPE: floppy disk COMPUTER: IBM PC compatible OPERATING SYSTEM: PC-DOS/MS-DOS SOFTWARE: PatentIn Release #1.0, Version #1.30 CURRENT APPLICATION DATA: APPLICATION NUMBER: US/08/485,355B FILING DATE: 07-Jun-1995 CLASSIFICATION: <unknown></p> <p>PRIOR APPLICATION DATA: APPLICATION NUMBER: US 08/440,522 FILING DATE: 12-MAY-1995 APPLICATION NUMBER: US 08/089,372 FILING DATE: 08-JUL-1993 APPLICATION NUMBER: AU PL4081/92 FILING DATE: 14-AUG-1992</p> <p>ATTORNEY/AGENT INFORMATION: NAME: Treccartin, Richard F. REGISTRATION NUMBER: 31,801 REFERENCE/DOCKET NUMBER: A-58631-2/RTT/DSS TELECOMMUNICATION INFORMATION: TELEPHONE: (415) 781-1989</p>				

:	TELEFAX: (415) 398-3249
:	TELEX: 910 277299
:	INFORMATION FOR SEQ ID NO: 40:
:	SEQUENCE CHARACTERISTICS:
:	LENGTH: 1704 amino acids
:	TYPE: amino acid
:	TOPOLOGY: linear
:	MOLECULE TYPE: protein
:	SEQUENCE DESCRIPTION: SEQ ID NO: 40:
:	US-08-485-355B-40
Query Match	5.7%; Score 515; DB 4; Length 1704;
Best Local Similarity	21.7%; Pred. No. 8.8e-36;
Matches 397; Conservative 179; Mismatches 553; Indels 700; Gaps 84;	
Oy	13 AAANSALANAVYV-----PFLSRVOTELLINIMOPROLYFREVL--WNHPI----Q 59
Dd	13 AAADAVYANVLOORAKVKDFAPPLALET--LHRITYP--LRFKGTLPPOHPILIAHQ 68
Oy	60 RVIHNELOYCARAGRCLEVGAHDRS---INDPNV-----LHRCFLRPVGDRVGRWSA 112
Dd	69 RVAEEVLNHFARGRS-TYLEIGPSLSALKHLGAPARNAYDYHG-C-TTKGTRDGSRHITA 126
Oy	113 PTRGPANCRRSALGLPP-----ADRTYCEDFSRCFAAETGYALSLLDL 160
Dd	127 LE-----SRSVAGRPEFKADASILANGIASRPFVCVDGVSCAFKRSGVTIANSLYDV 179
Oy	161 WPADVAEAMARGXRLVALHLPREVLLPCTYHTTSYLIIH-----DGD 206
Dd	180 TLEELANAFENHGLHMVRFMHMEPELLIDMNVNAELGYREHYIEEPMAVKDCAFQGD 239
Oy	207 -----RAVVTEGDTSGA--YNHDVSILRAMIRTTK 235
Dd	240 LRLHEPELDINESOERIERLERLAGSYSRRIVIFSGDDMDQDAYLHDPFHMLAVL---- 295
Oy	236 IYGDHP-----LVIERRAICGHVLLLTAPESPSPMPYPYPRSTEVYVASITGPCGS 289
Dd	296 LVRNYPPTPGFSLHTEVORRHGSSIELKITRAP-----PGDMILAIVPPTSGLCRI 347
Oy	290 PSLF--PSACSTKSTFFHAVPVHIMRMLFGATLDDQAFC--CSRILTLRGISYKVTVGA 346
Dd	348 PRIFYAAASGETEHKTIILTQHKVMMLNMQMOTREKEJLVMTVMSPARARLRALIVAS 407
Oy	347 LVANGSNVASEALTAXIITAAYLTICHORYLRTQAISGMKRLGYEHQOKFTTRLYSMLF 406
Dd	408 EYTESSMISPADLVRTVAVSLVLIHER-----RRAAVA----- 442
Oy	407 EKSGRDYIPGRLOFYAACRWLMSGFIHLDPRVLVFDESVPCCRCTFLKKVAGKRCCEMR 466
Dd	443 VKTADDV-----FGETSEWES-----LKHWLG-SCCCLR 471
Oy	467 WLGOECTCFLEPAEGLVGDHGDHNEAVEGSEVDPAEPALNDVSGTYAVAHGQLEALRYAL 526
Dd	472 NL-----KGTDV----- 478
Oy	527 NVPDIIARASRLTATVELVASPRDLBCRYUANKTFRTTYVVDGAHLEANGREOVYLSFD 586
Dd	479 -----VFKRVVADKYV----- 490
Oy	587 ASROSMGAGSHSLTYELTPAGIQVRISSNGLCDTATPPPGAPASMAPEVAFCALTRY 646
Dd	491 -----HSL-----CDI--ICD----- 499
Oy	647 NRFTQRHSLTGGLWLHREPGLLGIPPFPSPGHIMESANPCGBGLTYLTRTWSTSGFSDFS 706
Dd	500 -----VRLSPQOV-----GF----- 509
Oy	707 PPEAAPAMAATPGPHSTPRVSDIWVLPRPSEEROV-----DAAVPRAPDAGLPG 759
Dd	510 -----LPSVNPBAR--VFHDBRELVEYLREGACTINERYPST----- 543
Oy	760 PVVLTLPPEPPVHKRSLPPSRNRRLTYTPDGAKVYAGSLFEESDCDWLVNASNPGRHPRG 819

```

Db 544 -----PVPBEP-----QGDADL-W-----557
Qy 820 GGLCHAFYORPEAFYPTFEIMREGIAAYTLPRPIIHAVA---PDYREONPKRLAAY 876
Db 558 -----HATAASLPEYRA-----TLQAG 574
Qy 877 RETCSRGTAAPYLGSGIYQVPVSLSPAMERNRHPGDELYLTERAAMFPAANKPAQOV 936
Db 575 -----LNTDVKQKLTIL-----586
Qy 937 LTTEDTARTANLAEIDATEVGRACAGCTISP---GIYHOFAGVPGSGKRSI-----990
Db 587 -----ENAKT-----ID-----GLTSPVGLMEYE---GPPSGKGTITLIAL 623
Qy 991 -OOGDVVVVPTRELNSWRRR---GFAAFTPHPAARVTIGRR-----VVIDEA 1036
Db 624 EAAGKALYVAPTRELREAMDRIKPPSASATQVHA--LAILRATAEAGAFATVVIDEC 681
Qy 1037 PSLP--PHILLH--MORASSVHLIDGPNQIPATDFEHAGL--VPAIRP--ELAPTSMMXVTH 1091
Db 682 FMPPLVYVAIVHALESRSRIVLGDVHQIGFDQGSANMPLVRDYNVQCRRTFNQTK 741
Qy 1092 RCPADYV--CELIRGAVPKIQTTSRVLSLFMNEP---AIGOKLVYQAANKAN--PGAI 1143
Db 742 RCPADVATTFQOSLYPGCTTSGCVASISHVAPDRNSQAOTLCTQBEKSHGAEAM 801
Qy 1144 TVHEAGATFTETTTI--ATADARGLIOSSRAHAIVALTREKCVILDAAG--LIREVG 1199
Db 802 TVHEAGRTFAVILHYNSTAEOKLLAEKS--HLVGTIRHNLHYIRDPGIDIEROL 859
Qy 1200 ISDVIVNPNFLAGVEGXHRP-----SVIPRGNPDONIGTLQAPPSQI--SAHQOLA 1251
Db 860 HS-----AKAEFTDIPAPLEITTVKPESEVORN--EVMAITPQSATPHGAILHLR 909
Qy 1252 EELGHRP--APVAALVPRCPBELQGLYMPQELTVSDSV---LVEELTIVHCRMAAPSQ 1306
Db 910 KNMGPODCGVALAKTGEVEFGRAKINVELAEPDAPKPHAFQ--EGVQWVKVYNAASN 968
Qy 1307 KRAVUSTLVGRYGRRT---KLYEAHSDVR---ESLARPTPTGVOATTCELYELVEAM 1360
Db 969 KHAQALOTLLSRYSKRSADLPHE--AKEDYKRMILNSLRHMDMTVTEDBARAVFETOLKF 1027
Qy 1361 VKKGDSGSAVLELDCNRVSRITTFQKXCNKFTTGETTIAHGCVGGISAMSKTEPALG 1420
Db 1028 TORGGTEVDLEPDD--DPYTRIDIDFLMKTOOK--VSPKPINTGKVGGLIAHSLSLNFVLA 1084
Qy 1421 PWFRAIEKEIILALPNIFYG---DAYEESVFAAVSGASGCMVFENPSEFSDSTONNF 1476
Db 1085 AMIRILE--EILRTGSKTYVNSGLRDEEAMLEAKINOVPHATFVSADWTEFDTAHNNT 1143
Qy 1477 SLGLECVNMECGMPQMLIRLYLVHSAMILQAPKESLGFWMKKH-----SGEPSTLTM 1530
Db 1144 SELLEPAALLERIGTPAAAVNLFRERCGRKRLRA-----KGLGSVEVDGLDSSAAMTPCR 1198
Qy 1531 NTVMMKATIAHCEYERPDFAVAFKGGDSVYVLCSDYRQSNAAAALLIAGGGLK---LKVDYR 1587
Db 1199 NTFESAAMVLT--PRGVFAAFKGGDS--LLCGSHYLRDASRLHMGERYKTKHLKEVEO 1255
Qy 1588 PGLAGVAVVAGLGTLPVNVFEA---GR--LSE-----KNNCP-----1621
Db 1256 KIVPIYIGLIVSAQOYVL--DPVRSALKIFGRCTSELLISKIYEAVADITKMSDARYHSL 1314
Qy 1622 -----GPERAEQLRLAVCDFLRG 1639
Db 1315 LCHMSACYNVYAPESAAYIIDAIVVREGRG 1343

```

RESULT 8
 US-08-093-453B-2
 ; Sequence 2, Application US/08093453B
 ; Patent No. 5439814
 ; GENERAL INFORMATION:

```

APPLICANT: Frey, Teryl K.
APPLICANT: Dominguez, Geraldina
APPLICANT: Wang, Chin-Yen
TITLE OF INVENTION: Modified Infectious Rubella Virus
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Jamie L. Greene, Jones & Askew
STREET: 191 Peachtree Street, 37th Floor
CITY: Atlanta
STATE: Georgia
COUNTRY: United States
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50
COMPUTER: Macintosh
OPERATING SYSTEM: 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/093,453B
FILING DATE: 19 JUL 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 07/722,334
FILING DATE: 28 JUN 1991
ATTORNEY/AGENT INFORMATION:
NAME: Greene, Jamie L.
REGISTRATION NUMBER: 32,467
REFERENCE/DOCKET NUMBER: 07362-0101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404.818-3700
TELEFAX: 404.818-3799
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Rubella virus
STRAIN: Thierien
US-08-093-453B-2

Query Match 4.2% Score 374.5; DB 1; Length 2205;
Best Local Similarity 19.2% Pred. No. 3.2e-23;
Matches 454; Conservative 201; Mismatches 714; Indels 995; Gaps 104;

Qy 3 ITTAIDQALAAANSALANAVVVRPLSRVQTEILINLMQPROLVFRPEVLMNHPYRI 62
Db 40 VYTAAGKRAIV-----AVIPRPVFTQMVS-----DHP---AL 69
Qy 63 HNELEQYCARAGRCLEVGANPRS-----INDPNVLHRCLELPVGRDYORWYS-----111
Db 70 H-AISRYTRR--HWTEWG--PREALHVLIDPSBGL-----LREVARVERRWALCLHRT 118
Qy 112 -----APTRGPAN-----CRSALRGLP-----ADRTYC-FDGR 141
Db 119 AKRLATALAEITSEAMHADIYC---ALRGAPSGFFVYHEDVDVHGGRAYADRLCLYTPP 175
Qy 142 SRCFAETGVALYSLHDEMPADVA-----E 167
Db 176 QNCELMRTIDATLTLVAVDMLPVALAHVGDMDDLGIAMHLHDHDSGCPADCRGAGGPTP 235
Qy 168 AMARHCXTRLYAALHLPREYLLPFGTYHTTSTYLLIHDGBRAVVTYGGTISAGYNHDVSLI 227
Db 236 GYTRPCTFRIYQVL---PTAHNGRLYRCGPRMLTWRDCAVAELSWEAHCGHQ-----286
Qy 228 RAMIRFTKIVGDHPLVTEVRAIGCHFVLLLTAAPEPSMPAPYVPYPRSTRSEVYVRSIFGPG 287

```

287 -----ARVAVRCLPIRHVRSLOPS-----ARVR----- 311
288 GSPSLFSPACSTKSTFAVHVHMDRLMFGATLDQAFCCSRMTYLRGISYKVTGAL 347
312 -LBDVLHAEVGRKRWISLRPFVQRMISYCKTISPDAYISERFKKNALCHSITLACN 370
348 VANEGRN---ASEDALTAITAAVLTICHOBYLTQAIISGMRLGVEHAKFTRLYS- 403
371 VLOGMWGTCAEEDALCA-----YVAFRAMOSNARLAGIMKAKCAADSLSV 417
404 ---MLFEKSGRDYIPGRLOFYAOCRRMLSA-----GFHLDRIYVDESVCRCRTE 453
418 AGML-----DTI-----WDAIKRFLGSPVLAERMEKEODAAVAF----- 454
454 LKRVAGKCCFMRWLGEOCTCFLEPABGLVDGHNDNEAVEGSEVDAEPAHLDVSGTYA 513
455 -----RGPLEDGGRHLDYQ-----PKSPRPRIATW 484
514 VHGHOLE---ALIRALVPOD-----IAARAS 537
485 VHAASEDRHCACAPRCVPRERPSAPAGPDDEALIPMWLEARRALCREMPEALRAR 544
538 RLATFVELVASPDRLECRVTL-----GNKTFRTYVDGAHLEANGPEOY 581
545 ADTAAAP 604
582 VLSEFDSROSGAGSHSLTYELTPAGLO--VRISNGLDCTATFPFGAPSAAPGEVAAF 639
605 -----FAAGAHMCA---QARGLOAFVRY-----VPRPERMADGG--ARA 639
640 CSALYRNRTQORHSLTGMLHPEGLG-----IFPPSGHIESANPPCGESTLYRT 695
640 WAFEFRCACAAOR-----LLGEPYVHMLTYTGDVPOQL-----ALALRT 679
696 WSTSGFSDS-----PPEAAPAMAAT-----PGLP-----HS 724
680 LAOGGALALSVNDLPFGAIFDANAAYTAAYRAGROSNAASPPGDDPPRRARRSRHS 739
725 ---TPVSDIWLPPSESEFOVDAFVPPAPDAGLPGVVLTPPPPVHK----- 773
740 DARCTPPAP 787
774 ---PSIPEPSRNR-----RLLYTPDGAKYKASLSPESDOWL 808
788 VACEPSGPTSTRADPDSIDYESTARAGVYHLRVDIMDPPECKV-----V 835
809 VNASNPGRBGGLCHAFYORFPEAFYPTFIMREGLA--YTLTPRP----- 854
836 VMANEGLLAGSGVCAIFANATA-----LAANCRLAPCPTGEAVATPGHGC 884
855 ---IIHAVP-----DYVEONKRLAAVRE-----TCSRKGTAAVPLLGSIY--QV 898
885 GYTHIIHAVPARRBRDPRAALEEGALLERAYRSIVALAARRMCAVCAPLLGAGVYWSA 944
899 PVSLSFDAMERNRPGODELT-----TEPAAWFE 928
945 AEBLRAALAAATREVERVSLHICHPDRATLTHASVUGGLAARRVSPPTLELAS-CP 1003
929 ANKPAOPVLITTEDTATAMLALEIDATEVGRACAGTIS----- 969
1004 AGDPGRAGOSASAPPAITLG-----DATADEPRCCGCELCRTYRINDAYVNLMLERD 1058
970 -----PGIV-----HYOF----- 977
1059 RGATSWAMRIPEVYVYVPEHLATHFPLNHYSVLKPAEVRPPRCMGSGDMRCRGMHPQ 1118
978 -----TAGVPGSGKSRISIOGDVVDVVPTRRLN--SWRRGFAAFTPTTAA 1023
1119 VRCPTPSNAHALCRTCGRVPPASTRG--GELDPNTCWLRAAANAQAARAGVATSGACP 1175
1024 RVTIGR----- 1037
1176 KCAVGRALSEARTHEDEFAALSQRMSASHADSPDGTDPLDIMEITYGACSKRWVGSEH 1235

1038 SLPPHLLHLMORA-----SSVLL-----GDPNOI--PAI----- 1066
1236 EAPPDHLLVSLHRAPNRPMKGVLEVARRPBGPNPTGIFVCAVGGPRRVSDBRHLMLAVP 1295
1067 -----DFE-----HAGVPAIRPELAPTS--WXXYT-- 1090
1296 LSRGGTCAATDEGLAOAYVDLEVRRLGDDAMARALASVORPRKGPVNI RWNMAAGA 1355
1091 -----HRCPA-----DVCCELIR-----GAYPKI-- 1108
1356 GKTTIRLAATFREDLYCPCPNALLHEIQAKLRARDIDIKNAATYERLTPLAAYRIYI 1415
1109 -----QTSRVL-----RSLF--WNEP-- 1123
1416 DEAFITLGEYCARVASQTTAEVICVDROCGPHYANNCRTPVDDRWPTERSRHTWFPD 1475
1124 -----AIGKLYTQAAKANAAGAITVHE 1147
1476 CMAARLGLDYDIEGERTGFACNMDRGVDLHLAFSRETV--RRLHBAGIRAYTVRE 1533
1148 AOGATFETTTIATADARGL--IOSRAHAIVALTREKCVIID--APGLREYISDVIV 1205
1534 AOGMSVGTACIHYGRDGTVALALTRDLATVSLFRASDALYHLEBEDGSLRAAGLS----- 1589
1206 NNEFLAGEV-----GXHR-----PSVIRPNNDONLCTLAEPSCQISAYHOL 1250
1590 --AFLDGALAELEKEVPAGIDRVAVVQAPPLP--PADGIDPAQOVPPFCRTLELY 1644
1251 AEBLGRPAVAVALPPCPLEBGLIYMPBELVSDSVLYFELTDIYHGMAPSQRKAV 1310
1645 FGARGH--PHYADLNKVTGEHREVRVM--RISRLHLKNHTEM--PGTER--V 1689
1311 LSTLVGRYGRRTKLYEAAH--SDVRESLARIP--TIGPVQAT--CELYEL 1356
1690 LSAVCA-----VARYRAGEDSGLTRTAVARQHPRFPFQIIPPRYTAGVAGEMHTYLRER 1744
1357 YEAMVEKGQDSAVLEL-----DLGNRDYSRTTFQOKCNK-----T 1394
1745 IDLTDVYTMGVARELTDYRARRYPEIFAGMCTAOSLSVAPFLKATLKCVDALGPRDT 1804
1395 TGETIAHGKVGOGISANSKFCALFGFWFAIRKEITALLPPIFYGDVADEESVFNAAYS 1454
1805 EDCMAOGKAGLEIRAMAKMVOVMSPHRAIOKITIMRALRPOFLVAAGTEPEVDAMWQ 1864
1455 GAGSCWVENDFSEPDSTONFS-----LGLECYVMECGPOMLIRLYHLVRS 1503
1865 AHYTTNAIEVDFTPEFDMNQTLATRDVELFISAALLGPCA--ED-----YALRA 1912
1504 -ANILDAKRESLKGFMKKHSGEPGTLIMNTVMMATIAHCYER--DPRVAA--FRGDD 1557
1913 GSYCTJLRELSTGERTSGEPATLLHNT--VAMCMAMRMVPRKGVMMAGIFQGD 1967
1558 SVVLCSDYROSNAAL-----IAGCGLKLVDRPIGLVAGVAVPAGLCTLPDVR 1609
1968 MVLFPLEGARS--ALKKTPAEVGLFGPHIYKXNSTPSPSCG--HYGTAAGLFDVYMH 2023
1610 FAGRLSEKNMGPERAEROLAVCDFLRGLTNVAQCVDVVSRYGVSPGLVHNLIGML 1669
2024 QAIKYLCCRFE--DPDYLEEQVALLDLRLRG-----VYAAALP----- 2057
1670 QTTADGKAHFTETIKPYLDLTNSI 1693
2058 DTVAANAAAYDYSAERYLAIVREL 2081

RESULT 9
US-07-876-941A-31
; Sequence 31, Application US/07876941A
; Patent No. 5885768
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Bradley, Daniel W.

APPLICANT: Tam, Albert W.
 APPLICANT: Mitchell, Carl
 TITLE OF INVENTION: Hepatitis E Virus Peptide Antigen and
 TITLE OF INVENTION: Antidodies
 NUMBER OF SEQUENCES: 76
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dehlinger & Associates
 STREET: 350 Cambridge Avenue, Suite 250
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/876,941A
 FILING DATE: 01-MAY-1992
 CLASSIFICATION: 433
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 822,335
 FILING DATE: 17-JAN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 505,888
 FILING DATE: 05-APRIL-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 420,921
 FILING DATE: 13-OCTOBER-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 367,486
 FILING DATE: 16-JUNE-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 336,672
 FILING DATE: 11-APRIL-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 208,997
 FILING DATE: 17-JUNE-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Sholtz, Charles K.
 REGISTRATION NUMBER: 38,615
 REFERENCE/DOCKET NUMBER: 4600-0093.33
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 324-0880
 TELEFAX: (415) 324-0960
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 78 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 IDEAL ISOLATE: Fig. 11, ORF 1, aa 1285-1362
 US-07-876-941A-31

```

RESULT 10
US-08-801-263A-2
: Sequence 2, Application US/08801263A
: Patent No. 5811407
: GENERAL INFORMATION:
: APPLICANT: Johnston, Robert E.
: APPLICANT: Davis, Nancy L.
: APPLICANT: Simpson, Dennis A.
: TITLE OF INVENTION: System for the In Vivo Delivery and
: TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
: STREET: 1211 East Morehead Street
: CITY: Charlotte
: STATE: No. 5811407th Carolina
: COUNTRY: USA
: ZIP: 28234
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/801,263A
: FILING DATE: 19-FEB-1997
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Sibley, Kenneth D.
: REGISTRATION NUMBER: 31,665
: REFERENCE/DOCKET NUMBER: 5470-147
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-420-2200
: TELEFAX: 919-881-3175
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2500 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-801-263A-2

```

```

Db 1709 TDISLMDSESGSLFSSFGSDNYRQVAVDVHVAOEPAPVPPRLKMARLAARMQ 1768
Qy 1081 LAPTSMWVTHRCPADVCELRGAYPKIOTSRVLSLFNNEPALIGOKLVYTOAKANP 1140
Db 1769 EEPFP-----PASTSSADESLHLSFDGVSISFGSLFDGEMA---RLAAQPPASTCP 1817
Qy 1141 -----GATVHE---AGATFETTTIAT---ADARGLIOSRAHALV-----AL 1179
Db 1818 TDVPMFSGSFDSEIELSRVTESEPVLFSGFEPGEVNSIISRSVSPPRKORRRR 1877
Qy 1180 TRHTEKCV-----ILDAPGLREVGISDVYVNNFPLAGGEVGHRSVPTRGNPDQ 1230
Db 1878 SRREYCLTGVGGIYFTSDTGPGLHOKKSVLQNOLEPTLERVLEIYAVLDTSKEEQ 1937
Qy 1231 NLGTLQAFPPSCQISAVHOLAEE-----LGHPRAPVAAVLPCC-----PELE 1272
Db 1938 LKLRQAMPFTEANKSRQSRKVENOKAITTERLLSGRLVNSAIDQECYITTPKRSYS 1997
Qy 1273 QGL--LYMPELTVS-----DSVLVELTD-----IVHC-----1299
Db 1998 SSVFANYSDPFVAVVNNYLHENYPTVASYQITDEVDAYLDVYACLDTATFCPAK 2057
Qy 1300 -----RMAFSPORAVIULVGRGRKLYEAHSDVRESLARIPITIG-----1344
Db 2058 LRSTPKRHEHRAFNIRSAVPSAMONTL--QNVLLAATKRNQNTQMBELPTLDSATNVE 2115
Qy 1345 -----PVQATTCELYELVAVVE--KGODGSVLE-----LDLCNRD 1379
Db 2116 CFRRYACNDEWEFEFARKPIRITT---EFVTAYVARLKGPKAALFAKTHNLVPLQEV 2171
Qy 1380 VSRITFEOKKCKNTTG--ETIAHGKYGOGISAMSKTFPCALFGWFRATIEKIIALLP 1437
Db 2172 MDRFVMDKRDVKVTPGKTHKEERKV-QVYQAAEPLATATLGLIHVELVRLVAVLLPN 2230
Qy 1438 I---FYGDAVEESVFAAASGAGSCWFEENDESEFSTONFSGLECVVMECGMPQWL 1494
Db 2231 IHTLFDMASAEFDALIHNFHQGD-PVLETDIASFDKSQODAMALTGIMLELDGVQPL 2299
Qy 1495 IRL---YHLVRSAMIIQAPRESLKGFWKKGSGEPTLLMNTVNMMAIIAHCYE--FRDF 1548
Db 2290 LDLECAFGELISSTHLPETGTRFKFGAMMK--SGMFLTFVNTLVNVYASVLEERLKT 2347
Qy 1549 RVAFKGGDSV--LCSDYRORSMAALIAAGCGJLTK-----VYRPLGLYAGVAVPGL 1601
Db 2348 KCAFAIDDNDLHGVSDKEMAEKATWL--NMENVKIIDAVIGERPPYFCGFILODSV 2404
Qy 1602 GLTLDVVRAGRLSE--KMGPCP---ERAEQLRLAYCD-----FLRGLTVAVQVAVV 1651
Db 2405 TS--TACRVADPLKRLKRLGKPLPADDEQEDRRALLDETKANFRVGITDTLAVAVAT 2462
Qy 1652 SRVYVSPGLVHNLGLQTLADGAHF 1679
Db 2463 YEVDNITRPVLL---ALRTEAQSRAV 2485

```

RESULT 11
US-09-102-248-2
Sequence 2, Application US/09102248
Patent No. 6008035

GENERAL INFORMATION:
APPLICANT: Johnston, Robert E.
APPLICANT: Davis, Nancy L.
APPLICANT: Simpson, Dennis A.
TITLE OF INVENTION: System for the In Vivo Delivery and
TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
STREET: 1211 East Morehead Street
CITY: Charlotte
STATE: No. 6008035th Carolina
COUNTRY: USA
ZIP: 28234

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,248
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/801,263
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sidley, Kenneth D.
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5470-147
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEO ID NO: 2
SEQUENCE CHARACTERISTICS:
LENGTH: 2500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-248-2

```

Query Match 2.64; Score 233; DB 3; Length 2500;

Best Local Similarity 20.44; Pred No. 1,1e-10;
Matches 238; Conservative 141; Mismatches 457; Indels 332; Gaps 51;

```

Qy 803 SDC--DWLVNANSNGHRRGGGLCHAFYQRPPEAFYPTFEIMREGLAAYTLTP-RPIIHAV 859
Db 1359 ADCGEAVVNAANLGRPGEGVCRAIYKRWPNSE--TDSATETGAKLVYCOGKVIHAV 1416
Qy 860 APDRVQNP-----KLEAAYRETCS-----RGTAAVPLLSGTY-----QVPVLS- 903
Db 1417 GPDRF--KHPEAALKLLQNNVNAHADLVNEHNKISVAIPLLSGTYAAGKDRLEVS LNC 1474
Qy 904 -----FDA-----W-----ERNHRGDELYLTPPA-----924
Db 1475 LTTALDRTDADVITYCLDKKKERIDAVLQKESVTELKDEMDLIDELVWIHPDSC LK 1534
Qy 925 -----NMFEANKPAQPVLTITTEDTARTAN-----LALIEDAATEVG 960
Db 1335 RKGSTTKGKLYSFEETGKTHQAKDMAELKVLFPNDQESNEQLCAVILLETMAIRE-- 1592
Qy 961 RACAGCTIS-----PGIVHYOFTAGVPSGSKRSIQGDV--DVVVVPTRELNS 1008
Db 1593 ---KCPVDHNPSSSPPTLPLCLCMYAMTPERVRLRSNNVKEVTYCSSTPLPKYKIKNV 1648
Qy 1009 WRRR--GFAAFTPTTARVYLTGRVVIDEAPSLPHL-----1043
Db 1649 QKVQCTKVVLFPHTPAFVPARKYIEAPQAPPAQAEAPGVAVPTPPADNTSLDV 1708
Qy 1044 ---LLHMQRAS-----SVHLGDPNQIPADIFENAGLVAIRPE 1080
Db 1709 TDISLMDSESGSLFSSFGSDNYRQVAVDVHVAOEPAPVPPRLKMARLAARMQ 1768
Qy 1081 LAPTSMWVTHRCPADVCELRGAYPKIOTSRVLSLFNNEPALIGOKLVYTOAKANP 1140
Db 1769 EEPFP-----PASTSSADESLHLSFDGVSISFGSLFDGEMA---RLAAQPPASTCP 1817
Qy 1141 -----GATVHE---AGATFETTTIAT---ADARGLIOSRAHALV-----AL 1179
Db 1818 TDVPMFSGSFDSEIELSRVTESEPVLFSGFEPGEVNSIISRSVSPPRKORRRR 1877
Qy 1180 TRHTEKCV-----ILDAPGLREVGISDVYVNNFPLAGGEVGHRSVPTRGNPDQ 1230
Db 1878 SRREYCLTGVGGIYFTSDTGPGLHOKKSVLQNOLEPTLERVLEIYAVLDTSKEEQ 1937
Qy 1231 NLGTLQAFPPSCQISAVHOLAEE-----LGHPRAPVAAVLPCC-----PELE 1272

```


Db 2394 IIDAVIGERPPYFCGGFIILODSVTS--TACRVADPLKRLFKLGRPLPADDEQDEDERRAL 2451
 QY 1634 CD-----FLRGITNAQVCVIVSRVYVSGPLVNLGLMLOTIADGKAHF 1679
 Db 2452 IDETRAMFRVGTITGLAVALVATTRYEDNITPVLL-----ALRTFAOSKRAF 2497

RESULT 14
 US-08-801-263A-5

; Sequence 5, Application US/08801263A
 ; Patent No. 5811407
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnston, Robert E.
 ; APPLICANT: Davis, Nancy L.
 ; APPLICANT: Simpson, Dennis A.
 ; TITLE OF INVENTION: System for the In Vivo Delivery and
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
 ; STREET: 1211 East Morehead Street
 ; CITY: Charlotte
 ; STATE: No. 5811407th Carolina
 ; COUNTRY: USA
 ; ZIP: 28234
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/801,263A
 ; FILING DATE: 19-FEB-1997
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Sibley, Kenneth D.
 ; REGISTRATION NUMBER: 31,665
 ; REFERENCE/DOCKET NUMBER: 5470-147
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-420-2200
 ; TELEFAX: 919-881-3175
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2517 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-801-263A-5

Query Match 2.4%; Score 212.5; DB 2; Length 2517;
 Best local similarity 19.7%; Pred. No. 7.3e-09;
 Matches 235; Conservative 141; Mismatches 454; Indels 361; Gaps 50;
 QY 803 SDC--DWLVNANPGRHPCGGLCHAFYQRFPAFYPTFEIRREGLAAYLTLP--RPITIAV 859
 Db 1339 ADCQEAHVYVNAANPLGRBEGCRAIYKRWNSF--TDSATETGTRAKLTVCQCGKATIAV 1416
 QY 860 ADDYVEQNP-----KRLAAYRETCS---RGTAAVPLLGSGIY-----QVPYSLS- 903
 Db 1417 GPDFR--KHPEALAKLTQNAHVAADLVNHNKISVAIPPLSTGIYAGKRLEVSLSNC 1474
 QY 904 -----TDA-----W-----ERHNRGDELYLEPPA----- 924
 Db 1475 LTTALDRDADVTIYCLDKRMERIDAVLQIKESYIELKDEMEIDDELWTHIPDSCILG 1534
 QY 925 -----NMFENKPAQPVLTTFEDTARTAN-----LALEIDATEWVG 960
 Db 1535 RKGFSYTKKXLYFEGTKFHOAKDMAELIKVLPFDSDNSNOLCAVILGETMEAIRF-- 1592
 QY 961 RACAGCTIS-----PGIVHQTAGVAGSGKSRISIOGDV--DVVVVPTRELANS 1008
 Db 1593 ----KCPVDHNPSSSPKTLPLCLCMYAMTPREVRHRLRSNNKVKVVCSSSTPLPKYKIKV 1648

QY 1009 WRR--GFAATPHTAARTITGRVVIDEAPSLPHL----- 1043
 Db 1649 QKVOCTKVVLFNPHPAFVAPARKYIEAPEOPAAPQAQAEAEVATPTPPADNTSLDV 1708
 QY 1044 --LLHMOPAS-----SVHLGDPNQ 1062
 Db 1709 TDISIDMEDSSRGSLSFSSRSGSDNITSMDSSSGSSLEIYDRQVYVADVAHVAQEPAP 1768
 QY 1063 IPADIFEHAGLVPAIRPELAPTSMXVYTHRCPADVCELLRGAYPKIQTTSRVLSLFWNE 1122
 Db 1769 VPPRLKMMARLAARMQEEPFP-----PASTSSADESLHLSFGVSGMSFGSLPDE 1820
 QY 1123 PAIGKIVYTOAKKANPALT-----VHE-AQATFTETITITADAR 1165
 Db 1821 WG-----ALAAQAPASTCPDVPMSFGSFSDETEELSRRVTESEPVLSFGSEFG 1871
 QY 1166 GLIOSSRAHAIVAL-----TRHTKCYIL-----DAPQLREVGISDVIYNN 1207
 Db 1872 EVNSTIISRSVVSFPFRKQRRRRSRRTXYLTVGCVIISTDTGPHLOMESVLOQLTE 1931
 QY 1208 FFLAGGEVGHRRPSYIPRGNPDNLGTLQAPPPSCQISAYHOLAE-----IGH 1256
 Db 1932 PTLERNVLEIRIYAPVLDTSKEPOLKLRVQMPTEANKSRYSQSRKVENOKAFTERLLSL 1991
 QY 1257 RAPVAVALPPC-----PELEQGL--LYMPOELTVS-----DSVLYFELTD-- 1295
 Db 1992 RLYNSATDQPECKYITPKPSYSSSVANYSDBKFAVACNNYLHENYPTVASYQITDDEY 2051
 QY 1296 -----IVHC-----RMAAPSRKAVLSTLVGRYGRRTKYEYA 1328
 Db 2052 DAYLQKVDSTVACLDQATTCPRAKLSYPRKHREYRAPNTSAVPSAMQNTL--QNTLIAAT 2109
 QY 1329 HSDVRESLARFTPTIG-----PVQATTCELYELVEAMVE- 1362
 Db 2110 KNCNVYQKRELPTLDSATFNVECFKRYACNDEYWEFARKPIRITTT-----EFTAYVAR 2165
 QY 1363 -KGODSAVLE-----LDLCNRDVSITTFQKXCNFTTG--ETIAHGKVGQGISMSMT 1414
 Db 2166 LKGPRAALFAKTHNLVPLQEVPMDFVDMKRDVAVTGTGHTTERPV--OYLOAEPL 2224
 QY 1415 PCALFEPNFAIEKEILLALPPNI---FYGDAYEESVFAAASGASGVFENDESEFDS 1471
 Db 2225 ATAVLCGIRHRELVRLTAVLLPNIHTLFLDMSAEDFPAIIAHEHKQGD--PYLEDIASFPK 2283
 QY 1472 TQNNFSLGLECVAMEECGPQMLIRL---YHLVNSAMLLQAPKESLKGFMKKHSEPGT 1527
 Db 2284 SODDAMALTGMLIEDLVGDPLDLIECAFGEISSTHLPTRFRKFGAMMK--SGMPLT 2341
 QY 1528 LLMNTVNMMAIIAHCE--FRDFRVAAPFKGDDSVY--LCSDRQSNMAAALIAAGCKLIX 1583
 Db 2342 LTVNTVLNVIASRYLEERLTKSKCAFTGDNITHGVVSDKEMARCATWL--NMEVK 2398
 QY 1584 ----VDYRPIGLYAGVVAVAPGIGTLPDVVRFAGRLSE--KMWGPP---ERAQOLRLAV 1633
 Db 2399 IIDAVIGERPPYFCGGFIILODSVTS--TACRVADPLKRLFKLGRPLPADDEQDEDERRAL 2456
 QY 1634 CD-----FLRGITNAQVCVIVSRVYVSGPLVNLGLMLOTIADGKAHF 1679
 Db 2457 IDETRAMFRVGTITGLAVALVATTRYEDNITPVLL-----ALRTFAOSKRAF 2502

RESULT 15
 US-09-102-248-5
 ; Sequence 5, Application US/09102248
 ; Patent No. 6008035
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnston, Robert E.
 ; APPLICANT: Davis, Nancy L.
 ; APPLICANT: Simpson, Dennis A.
 ; TITLE OF INVENTION: System for the In Vivo Delivery and
 ; NUMBER OF SEQUENCES: 12


```

OY      1621  PGPERRAEQRLAVCDFELGRTNVAQVCVDV/SRRVGVSPGLVHNLIGMQLTADCKAHT 1680
Db      1621  PGPERAEQIRLAVCDFFILGLTNVAGVCVDVSRVGVSPGLVHNLIGMLQTADGKAHT 1680
OY      1681  ETIKRPVLDTNSIIORVE 1698
          |||||||
Db      1681  etikrpvldtlnsiigrve 1698

RESULT      2
ID          W93408
W93408 standard; Protein; 1708 AA.
XX
AC          W93408:
XX
DT          11-JUN-1999 (first entry)
DE
DE          Swine HEV ORF 1 protein.
XX
KM          Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
KW          vaccine; immunise; infection; detection; diagnosis; prevention.
XX
OS          Hepatitis E virus.
XX
PN          MO9904029-AZ.
PD          28-JAN-1999.
PF          17-JUL-1998; 98WO-US14665.
PR          18-JUL-1997; 97US-0053069.
PA          (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI          Emerson SU, Meng X, Purcell RH;
DR          MPI, 1999-132270/11.
XX
PT          New isolated swine hepatitis E virus - used to develop products for
PT          the diagnosis, prevention and treatment of hepatitis E virus
PT          infection in mammals, particularly humans
PS          Example 1; Fig 6D-J; 70pp; English.
XX
CC          This invention describes a swine hepatitis E virus (HEV) and its natural
CC          mutants which are capable of cross-reacting with antibodies reactive
CC          with a human HEV strain or natural mutants. The HEV and the proteins
CC          can be used in vaccines for immunising against HEV infection. The swine
CC          HEV can be used in humans to prevent possible infection by human HEV. The
CC          swine HEV can also be used as a therapeutic treatment for infection by
CC          other strains of HEV. The swine HEV can also be used for the production
CC          of antibodies which can be used in therapy, detection and diagnosis. The
CC          products can also be used for determining the susceptibility of cells or
CC          organs to infection with swine HEV. The swine HEV is particularly useful
CC          for the development of agents for the prevention, treatment and detection
CC          of human HEV because it is not a human virus and thus can be handled both
CC          experimentally and clinically without fear of severe infection and/or
CC          contamination.
SQ          Sequence 1708 AA;

Query Match      97.8%; Score 8812.5; DB 20; Length 1708;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1661; Conservative 7; Mismatches 30; Indels 1; Gaps 1

OY      1  PGTTAIEQAAALAAANSALANAVVPFPSPROTLLINIMOPROLVFPPEVLANHPRIOR 60
          |||||||
Db      10  pttaitqaaalaanaalanavvrpfllstrgtelllhmprdqvlfrpeylmnpdqr 69
          |||||||
OY      61  VIHNLEQYCARAGCDEVGAHPRSINDNPVLAHCFLRPVGKDVORWSAPFRGPAAN 120
          |||||||

```

D	b	70	aimeleqyrcratagoclewygauprfrindhpnvlnhcfllrpgrivqcywysaerpcrgaan	128
Q	y	121	CRNSALRGLEPPARARYCFDGFSCFAEATGYALYSLHDLMPADVAEMARHGXTLYAA	180
D	b	130	crsalrglppvdrtycfdfgsrcafaetgalyalslndlpwadaeamaahgmrtlyaa	189
Q	y	181	LHLPEVLEPRGYHNTSYLLIHDBGRAVVTGEGTSAGVYNDVSLRAMTRTKYGDH	240
D	b	190	lhlpevlppgyhtntsylllhndgravvtvegdsagymndvslrawlrfcklygdh	249
Q	y	241	PLVIERRAIGCHFVLTLTAAPESPMPVVPYPRSTEVYVRSIFGPGSPSLFSPASCTK	300
D	b	250	plvierrraigchfvllltaapepsmpvpyprstevyvsrlfpgpspslfpasctk	309
Q	y	301	STFHAVPVHIMDRMLFGATLDQAFCCSRMLTMYLKGISKYVTGALVANEGMNASDEAL	360
D	b	310	stfhavpvhlwdrlmldfgatlddqafccsrmltmylrglskyvtgalyanegmasedal	369
Q	y	361	TAHTTAAYLTLCHQRYLRQALISKMRRLGVEHAQKFTRLXSWLFEEKSGDYIPGRQD	420
D	b	370	tavtaaylthchqyrlrtqalskmrlevehaqkfltrlyswlfeksgdyipgrdq	429
Q	y	421	FYACCRMLNSAGFLHEDRVLVDFESPCKCRFTFLKKVAKFCCFMRMLGOEOTCLEPAE	480
D	b	430	fyaqccrmlnsagflhdrovrlvdfespcrctrlkkvqakfccfmrvlygetctclpeae	489
Q	y	481	GLVGDHGDHDEAYEGSEVDEAPBAHLDVSGTAVYAHGHOLEALYRALNVPDIAARASRLT	540
D	b	490	glvgdghdhdneayegsevdapbahldvsgtyavbhgrlealylralnvpdhlaarasrlt	549
Q	y	541	ATVELVASPDRLCERYLGNKTRFTTYVDCAHLEANGPQOYLSFDASQSGAGSHLT	600
D	b	550	atvelvaspdrlcertylgnktrfttyvdcahlengpqqoylsfdasqsgagshlt	609
Q	y	601	YELPRAGLOVSISSNGDLCATFPFPGGAPAAAGEEVAACSLYRNYRFTOSHSLTGCLM	660
D	b	610	yelpraglovisssngldcatfppggaapaagevaacsalrynyrftqhsltgclmw	669
Q	y	661	LHPEGLLGFPPFSPGHIWESANPFCGEGTLYRRTWSTSGFSSDSPPEAAPAMAATPG	720
D	b	670	lhpegllgflfppfspghilwepanpfcgegflylrrtwstsgfssdsppeaapvlaaapg	729
Q	y	721	LPHSTPPVSDIWVLPRPSEEFQVDAPVPPAPDAPGLPCGPVULT-P-PPPPPHKRSIPPP	779
D	b	730	lphstppvswvlprpssesqydaasvppapepajlpsivlclpprlppvkrxpppp	789
Q	y	780	SRNRLLTYTYDDGAKVYAGSLFESDCDMVLNANPCHPRGGGICHAFYORPEATYPIEF	839
D	b	790	srllrlytydgakvyagslfesdcmlwnasnpnrppgajchaigyrfpeatypief	849
Q	y	840	IMREGIAVYTLTPPRLIHAAVAPDVREOVQNKRLLEAAVRETCSRNGTAAPPLILGSGITQVP	899
D	b	850	lmreglaayltlpprlhavaapdyrvegnpkrlleaayretcsrgtaaypllgsllyqvp	909
Q	y	900	VLSLFDANERNHRPGDELYLTERPANKFEANKPQAPVLTTEDTARTANLALEIDATEV	959
D	b	910	vlsldawernhrpgdeylterpaafeanpkaqpalttledartanlaaleidaatdv	969
Q	y	960	GRACAGCTISGCIYHYQFTAGVRESGSKSRITQOGVDVYVPPRRELRNWRRRGAAATPP	1019
D	b	970	gracagctclspglvhyqftagvpsgsksrslggdvdvvyvprcltnsmwrrrgtaatpp	1029
Q	y	1020	HPTAAVTTGRRVNVIDEAPSLPRLPHLLLHHMRASSVHLDBDPNDIPALIDEHAGLVPAIRP	1079
D	b	1030	htaavvtlgrvrvideapslphllllhmrassvhljqdnpqipaidehaglvpaairp	1089
Q	y	1080	ELAPTSMMXVTHRCRADVCELIRGAYPKIOTTSRVLSLFWMNEPAIGOKLYVXTOAKAAN	1139
D	b	1090	elaptsmwvthrcpadvcelirgaypkigttsrvlslfwmepalgklyftvgakaan	1149
Q	y	1140	PGATVVEHAGACATTEETTIIATADAKLIDSSRAHAIVALTRKTECVTLDAPELLREVG	1199
D	b	1150	pgatvvehagatfctetlilatadarglqssstahavaltcnktecvllldapqllirevg	1209

QY	1200	ISDVIVNNFPLAGEGVXKHRPSVYPRKNRPQNIQCTIQAPRPSQISATYHQIALEELGHRPA	1235
Db	1210	ISdviwnffllaggevgvhhprpsvlyprgnpqnlgltlqatfppscqisayhqlaeelghpra	1269
QY	1260	PVAANLPPCPPELEOGLLYMPDELVSQSVLYFELTQIVHCQMAAPSGQKAVLSTLYGRY	1319
Db	1270	pyaavllppcpelqeglllympelvtvsdsvlyvfeltqivncmaapsqtrkavlstlygrY	1329
QY	1320	RRTKLYEAHSDVRESLARFIPRTIGPVQATTCELYELVAMYEKGQDGSVLELDLCNRD	1379
Db	1330	rtklyeaahsdvreslarfiprtlqprvatctcelylevamwekgqgsavyleldlcnd	1389
QY	1380	VSRTTFPPKXKXKNTTGTGTTLAHGKVGCGISANMKTGCALFGFMFRAIEKEILLPPNIF	1439
Db	1390	vsrttfqkcnkktlgteltlahgkvvggisawskticalfgpwfraiekeliallppnlf	1449
QY	1440	YGDAAEEVSFAAANAAGSGSCGVFENDSEFDSPTQNNFSLGLECYVMEECGMPQWILRLYH	1499
Db	1450	ygdaveesvfaaavsgsgscmvfendsefdsqnnfslglecyvmeecgmpqwlrlYh	1509
QY	1500	LYRSAMWIIQAPKESLKGPMKKHSGEGCTLLMNTVMMAITAHCHYERDFRVAARFGDDSV	1559
Db	1510	lyrsamllqapkesllkgfwkhsgepgctllmnlvmmaiahcyefdrvaafkgddsv	1569
QY	1560	VLCSPYRQSRNAALNLINGCGIKLKVDYRPIGLAGVYVAPRGITLPDYVRFAGRLSEKNW	1619
Db	1570	vlcspdyqrsmnaaalnlagcgiklkvdyrpi9lyagvvpaprgitlpdyvrtfagrlseknw	1629
QY	1620	GGPERAEQRLAACDRLRGITLTVNAQVCYDVVSRYKVGVSFGLVHNLIGMLQTIADKAHF	1679
Db	1630	gpgperaeqrlrlavcdfirglrltnvaqcvdvsrykvgvsfglvnhlignmlqtiaadgkafh	1689
QY	1680	TEITIKPVLDLTNSTIQRVE 1698	
Db	1690	teitkpvldltlnstlqrve 1708	
RESULT	3		
VJ1384			
ID	VJ1384	standard; Protein: 1708 AA.	
XX			
XX	VJ1384;		
AC			
XX			
DT	12-Oct-1999	(first entry)	
XX			
DE	HEV-US2 ORF1 protein.		
XX			
KW	Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;		
KX	vaccine; passive immunisation.		
XX			
OS	Hepatitis E virus.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference 322	/label= unknown	
FT		/note= "encoded by C8c"	
FT			
FT	Misc-difference 331	/label= unknown	
FT		/note= "encoded by RAc"	
FT			
FT	Misc-difference 445	/label= unknown	
FT		/note= "encoded by RAc"	
FT			
FT	Misc-difference 448	/label= unknown	
FT		/note= "encoded by RAc"	
FT			
FT	Misc-difference 634	/label= unknown	
FT		/note= "encoded by RTG"	
FT			
FT	Misc-difference 646	/label= unknown	
FT		/note= "encoded by CCY"	
FT			
FT	Misc-difference 811	/label= "encoded by GCS"	
FT			

FT	/label= unknown	
FT	/note= "encoded by TKT"	
FT	Misc-difference 1533	
FT	/label= unknown	
FT	/note= "encoded by TAY"	
FT	Misc-difference 1578	
FT	/label= unknown	
FT	/note= "encoded by RGC"	
FT	Misc-difference 1691	
FT	/label= unknown	
FT	/note= "encoded by GAR"	
PN	W09919732-A1.	
PD	22-APR-1999.	
XX		
XX	15-OCT-1996; 98MO-US21941.	
PF		
XX	15-OCT-1997; 97US-0061199.	
PR		
XX		
PA	(ABBO) ABBOTT LAB.	
PI		
XX	Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;	
DR	WPI: 1999-288017/24.	
XX		
XX	N-PSDB: 200267.	
PT		
XX		
PS	Claim 6; Page 216-221; 260pp; English.	
XX		
CC	The invention provides a method for detecting a US (sub)type hepatitis E	
CC	virus (US-HEV), or its naturally occurring variants in a sample by	
CC	treatment with a binding partner specific for a marker of the virus, and	
CC	then detecting any complex formed. The method is used to diagnose	
CC	infection with US-HEV. Polypeptides from US-HEV, antibodies specific for	
CC	open reading frames (ORF) in US-HEV and host cells expressing these ORFs	
CC	are useful in vaccines or for passive immunisation. The polypeptides are	
CC	also used to raise specific antibodies (useful as immunoassay reagents).	
CC	Fragments of nucleic acid from US-HEV are useful as primers and probes in	
CC	usual hybridisation and amplification assays for detecting infection. The	
CC	present sequence represents a HEV-US2 ORF1 protein.	
XX		
SQ	Sequence 1708 AA:	
	Query Match 97.1%; Score 8753.5; DB 20; Length 1708;	
	Best Local Similarity 97.2%; Pred. No. 0;	
	Matches 1651; Conservative 8; Mismatches 39; Indels 1; Gaps 1	
QY	1 PGIITTAIEGALAAANSALANAVVVRPELSRVOTETILINLMOPROLVREPEVLWNPOR 60	
DB	10 PGIITTAIEGALAAANSALANAVVVRPELSRVOTETILINLMOPROLVREPEVLWNPOR 69	
QY	61 VTHNELEQVRCARACLEVGGAHPSSINPNVWLHHCFCRPVGRDQVRYSAVTRGPAN 120	
DB	70 VTHNELEQVRCARACLEVGGAHPSSINPNVWLHHCFCRPVGRDQVRYSAVTRGPAN 129	
QY	121 CRRSALRGILPADRYTCFDFGSRCAFAETGVALYSLHDLWPAADVAAEMARRGXTRLVAA 180	
DB	130 CRRSALRGILPADRYTCFDFGSRCAFAETGVALYSLHDLWPAADVAAEMARRGXTRLVAA 189	
QY	181 LHLPEVILLPECTYTTSTGLLTHDDRAVVTYEGGTSAGYNNIDVSTLRWITTKIVGCH 240	
DB	190 LHLPEVILLPECTYTTSTGLLTHDDRAVVTYEGGTSAGYNNIDVSTLRWITTKIVGCH 249	
QY	241 PLVERVAIGCHFEVLLTLTAPESPMPVPRVPRSTEVVRSIFPGSGSPSLFPACSTK 300	
DB	250 PLVERVAIGCHFEVLLTLTAPESPMPVPRVPRSTEVVRSIFPGSGSPSLFPACSTK 309	
QY	301 STEHAVPVHIMDRMLFGATLDDQAFCCSRMLTMYLRGISTYKVTVGALVANEGWNASDAL 360	
DB	310 STEHAVPVHIMDRMLFGATLDDQAFCCSRMLTMYLRGISTYKVTVGALVANEGWNASDAL 369	

QY 361 FAXITAAVLTICHORYLROAISKGRRLGVEHAOKETITRLYSWLFPEKSGRDYIPGRLO 420
 Db 370 taviiaaylitichqyrlitqaiskgmtilevenahqkfiltrlyswlfleksrdy1ipqr1q 429
 QY 421 FYAQRMRSLAGFHLBPRLVLPDESVPICRQRTFLKRVAGKFCCEMRVLGECCEFLPAE 480
 Db 430 fygcrtrrtaagfhlkprxlrvfdesvpcrcrtflkkrvagkfccefmwlgceccflepa 489
 QY 481 GLVGDHGHNEAVEGSEVDPAPRAHLDVSTYAVHGHOLEALTRALNVPDIDARASRLT 540
 Db 490 g1vgdgghneavegsevpaprahdvsg1tyavhghlealtralnvpdidaarasrlt 549
 QY 541 AATVELASPDRLCECTVVLGKTRFTYVODGAHLEANGPEQVLSFPASRQSMAGSHLT 600
 Db 550 atvelaspdrlcectvvlgnktrltcvtvdgaahleangpeevlstafrsqsmagshlt 609
 QY 601 YELTPAGLVRISSNGLDCAATPPPGAPSAAGAEVAFCALRYNRFTQRHSLNGLW 660
 Db 610 yeltplaglvrkissngldcatfpagapaaagexafsalrynrftqrhslnglw 669
 QY 661 LHPEGLLGIPTPPSPGHIMESANPFCGEGTLTYRTWSTSGSSDFSPPEAAPAMATPG 720
 Db 670 lhpegllg1tppspghimesanpfcgeg1tyrtwstsgssdfsppeaapasaag 729
 QY 721 LPHSTPPVSDIWLPPPSSEFQVDAPVPAPAPAGLPGVULT-PPPPPPVHKPSITPP 779
 Db 730 lpytpvpsdlw1lpppseshvaasvpsvpeag1tsp1lppppppvkhpsitpp 789
 QY 780 SRMRRLLYTPPDGAKYAGSLFESDCDMLVNASNPGHRCGGCHAFYQGFPAFPTPE 839
 Db 790 p1rr1llytpdgakyyagslxesdcwlvnasnpg1hrcgg1chafyqgfpaefstet 849
 QY 840 IMREGLAAYTLTRPPIIHAAVPDYRVEQONKRLAAVRETCSRRAAYPLLSG1YQVP 899
 Db 850 imreglaaytltrp1ihavapyrveqnpkrlaaeyretcsrrgaaypl1sg1yqvp 909
 QY 900 VLS1SPDAMENNRHRCDELITTEPAANMFENAKRAQPVLTTEBTARANALALDIAATEV 959
 Db 910 vls1stpdamenhrhrcdel1y1tepaanmfenakpqpallttdetartaalalidaatev 969
 QY 960 GRACACCTISPG1VHYOFTRGVPGSGKRSIQGDVDVVPVPRTELNSMRRGFAATP 1019
 Db 970 gracagctispg1vhyoftrgvgpgsgksrsiqgddv0vvvprtelnsmrtrrgfaa1p 1029
 QY 1020 HFAARVTIGRRVVIDEAPSLPRLHLLHMQRASSVHLGDPNOIPALDFEAGLVPAIRP 1079
 Db 1030 htaarvtigrvv1deaps1p1hll1hmqraassvhl1gdpnq1paldfehag1vpa1rp 1089
 QY 1080 ELAP1SMWXYTHRCPADVCELRGATPKIOTSTRVLSLFWNEPAIGOKLVX1QAAKAA 1139
 Db 1090 elap1stwmxythrcpadvcel1rgaypk1q1tsrv1st1wnepa1gqk1v1fcqaakaan 1149
 QY 1140 PGATVTHEAGATFETTTTATADARGLIOSSRAHAIVALTTRTEKCVILADAGLREVG 1199
 Db 1150 pga1tvtheagatfett1tatadarg1llqssrahaiva1ltnrtekv1l1dapg1l1revg 1209
 QY 1200 ISDVIYNNEFLAGGEVGHXRPV1PRGNPDONLGTLOAPPSQOISAYHOLAELGHRPA 1259
 Db 1210 isdviynneflaggevgvghrpsv1prgnpdn1gt1qafppscq1sayhqlaeelghrpa 1269
 QY 1260 PVAANV1PRCELEOGILLMYQOELTVSDSVLFFELTDVHCRMAAPSGKRAV1STLVGRYK 1319
 Db 1270 pvaav1prpceleg1llmyqel1vds1v1f1td1v1hcrmaapsrkav1st1lvgr1y 1329
 QY 1320 RRTKLEAAHSDVRES1AR1PTIGVQATTCLELVEAMVEKGGDSVLE1DLCNRD 1379
 Db 1330 rrtk1yleaahsdvres1ar1pt1gvr1atccel1yelveamvekggds1v1eld1cnrd 1389
 QY 1380 VSRITFROKKCNKFTTGET1AHGRVGGISAMSKTFCALFGPWRALKEK1TALLPNNIF 1439
 Db 1390 vsr1tffrokkcnkfttget1ahgkvvg1sawsktfcalfgpwfrakek1allpnnif 1449

QY 1440 YGDATSEVFAAASGAGSCMVFENDESPDSTONNFSGLGECVMEGCMQWILRLYH 1499
 Db 1450 ygdatee1vfaaasga1scm1vfend1se1ds1qnm1fs1gl1ecv1mee1c1mpq1l1r1yh 1509
 QY 1500 LVRSAM1LOAPKESLKGFMKKHSGEPGTL1MNTVWMA1IAHCYERDFRVAARFKGDDSV 1559
 Db 1510 lvrsav1llqpk1es1kgf1wkkhsg1epg1tl1w1v1wma1ia1hcx1ef1df1vaaf1kg1dsv 1569
 QY 1560 VLCSVDYRQSNAAAL1AGCCLAKVDYRPIGLYAGVVAAGL1TDPVYVFAGR1SEKKN 1619
 Db 1570 vlcsdyr1q1x1naa1la1agc1kl1kvdyr1p1gl1y1agv1v1ap1gl1tp1dv1v1f1agr1se1kn 1629
 QY 1620 GPGPERAEOLRLAVCDPLRGLTNVAQCVDVSVKVGSPGLVHNL1GM1QTTADGKAH 1679
 Db 1630 gpgperaeq1rl1avcd1l1r1gl1tnv1aqcvdvsv1kv1gsp1gl1v1hnl1gm1q1t1ad1g1k1ah 1689
 QY 1680 TETIKPVLDTNLS1IQRVE 1698
 Db 1690 t1x1kp1v1d1t1n1s1i1qrve 1708
 RESULT 4
 R91813
 ID R91813 standard; Protein; 1693 AA.
 AC R91813;
 DT 26-NOV-1996 (first entry)
 DE Hepatitis E virus strain SAR-55 ORF-1.
 XX
 DE Hepatitis E virus; SAR-55 strain; enteric transmission;
 KW Hepatitis E virus; antigen; detection; antibody; vaccine;
 KW Immunisation; infection.
 XX
 OS Hepatitis E virus.
 XX
 FH key Location/Qualifiers
 FT Misc-difference 1238 /note- "corresponding codon CAG"
 FT Misc-difference 1455..1693 /note- "10 bp nucleic acid sequence TGGTNTTGYCA has to be inserted between nucleotides 4390..4391 of T27394 before these amino acid residues can be decoded"
 FT
 FT
 FT
 W09610580-AZ.
 PD 11-APR-1996.
 XX
 PF 03-OCT-1995; 95WO-US13102.
 XX
 PR 03-OCT-1994; 94US-0316765.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Emerson SU, Purcell RH, Tsarev SA.
 XX
 DR WPI: 1996-209320/21.
 XX
 DR N-PSDB: T27394.
 XX
 PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
 PT antigenic protein useful in diagnosis, prophylaxis and treatment of
 PT hepatitis E virus infection
 XX
 PS Disclosure; Pages 9-13; 121pp; English.
 CC
 CC The present sequence is the protein prod. of ORF-1 from the
 CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
 CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
 CC protein encoded by the structural region of the virus (i.e. ORF-2),
 CC which is capable of forming HEV like particles, is useful for the
 CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,

CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
 CC protein, and anti-HEV antibodies generated using the protein, can
 CC also be used in vaccines for immunising an animal against HEV
 CC infection. The protein is identified as a band of greater than
 CC 50 kD following SDS-PAGE of cell lysates of insect cells infected
 CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
 CC expression vectors pPIC9-1779, -1780 and -1781.

XX Sequence 1693 AA;

Query Match 82.7%; Score 7456; DB 17; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1400; Conservative 110; Mismatches 160; Indels 42; Gaps 7;

QY 1 PGITTAIBQALAAANSALANAVVVRPELSRYQTEILINLMQPROLYRPEVLMNPIOR 60
 DB 10 pgitatqgaalaansalanavvrpfisqgqlellnlnmqprqlvtrpevfwhpqr 69
 QY 61 VIHNELEOYCARACRCLEVGANHPISINDPNVLRHCLRPVGRDVGQRYKSAPTRGPAAN 120
 DB 70 vihnleleycraargrcleigahprslndpnvhrclrpgrdvgrwytaftrgpaan 129
 QY 121 CRRSALRGLPADRTYCDGFSRCAFAETGYALVSLHDLMPADVAEAMARHGXRLYAA 180
 DB 130 crrsalrlgpadrtycdgfsrcafaetgyalvslhdmppsdaeamftrgmrlryaa 189
 QY 181 LHLPEVLLPGTYHTTSYVLLIHGDRAVVVTEGDSAGVNHDSILRAMIRTKTKVDH 240
 DB 190 lhlpevllpgtytaayllihdgrvvvteygdtsagynhdvsnlswitrtkvdh 249
 QY 241 PLVIERVAIGCHFYLLTAAPESPMPYVPRSTEVYVNSIFPGGSPSLFPACSTK 300
 DB 250 plviervaligchfylltaapepmpypvprstevyvislfpggslpslfpstcsk 309
 QY 301 STFNAPVPHIMDLMLFGATLDDQAFCCSRMLTFRGTSYKVTGVALVANEQNMNSDAL 360
 DB 310 stfnavphimdlmlfgatlddqqfcsrlmlylrgsykvtgvalvaneqnmnsedal 369
 QY 361 TAXITTAALVTCOHQRYLRTQALSKGMRLGVEHAKFTIRLYSWLFEKSGRDIYIPGRLO 420
 DB 370 taxitaaylvtcqhrylrtqaliskgmrlerehaqftrlyswlfeksgrdyipgrle 429
 QY 421 FYACORRLASGFHLDPRVLYFEDSVPRCRTLKLVAGKRCFCFMRWLGQECTCLEPAE 480
 DB 430 fyaqrrtlasgfhlprvlyfdesapchrtalrkavskfctfkmklygqectcllqpa 489
 QY 481 GLVGDHGDNDNAYEGSEVDPAEPALHDVSGTYAVHGHOLEALRYALNYPDIIAARSRLT 540
 DB 490 glvgdqgdndnayegevdpaesaidsysgyvvpqtalqplqgalpaelvaaragrlt 549
 QY 541 ATVELVASPDRLECRVYLGKNTFRTTVVDGAHLEANGPEQYVLSFDASRQSGAGSHSLT 600
 DB 550 atvsvsqvdrlecdvlgkntftrtvdgahleangperhmsfdasqsgmaagpfslt 609
 QY 601 YELTPAGLOVATISSGLDCTATFPPGAPSAAPGEVAAFCALRYKNFTORHSLTGLWM 660
 DB 610 yaasaaglevyvaagldhravfagysprzargevtafcaalyfrneaqrlsltgntw 669
 QY 661 LHPGGLGIFPPFSGHIMESANPFCGGLTYTRTW-----TSGESSDFSPPE 709
 DB 670 fhpeglliprlaprlspghwesaprcgescllyrtwsevdavspapqdlgfts-----e 724
 QY 710 AAAPAMATPGLPHSTPEVSDIWLPRPSEEFQYDAAV-PPAPDPA-GLPGPVLLPFP 767
 DB 725 psipstratp-----tp-----aaipppadpdpstlsapargepap 761
 QY 768 PPVYHAKSIP-PPSRNRRLTYTPDGAKVAGSLFESDCDMLVNASNGHRRGGGLCHAF 826
 DB 762 gataarapaltqtatrrllftypdgskvfagslfestctvlnasnvdhprgglchaf 821
 QY 827 YORPEAFYPTFEFIMRGSLAYVTLTPRIITAAVAPDRVEONPKRLAAYETCSRKETA 886
 DB 827 yorpeafyptfefimrgslayvltlprriitaaavpdrveonpkrlaayetcsrketa 886

DB 822 YGRYPASIDAESFYMDGAAAYTLPRPILHAVAPDYRLHNPXKLEAAAYETCSRLYGLA 881
 QY 887 AYPILGSGIYQVPSLSPDAMERNHRPDELYLTPPAANWFEANPAPQVLTTFEDTRT 946
 DB 882 aypilgsgiyvpslspdamernhrpdeyltppaanwfeanpaptqvlttfedvart 941
 QY 947 ANLALEIDATEVGRACAGCTISPGIVHYOFTAGVPSGSKSRISIOGDPDVVVPTRRLR 1006
 DB 942 anlaleidsatdvgracagcvtlpgvvqyftagvpsgsksrsltqadvdvvpvtrlelr 1001
 QY 1007 NSMRRGGAATPHTPAAAVTIGRRVVIDEAPSLPHLLLLHMQRASSVHLLGDPQIPAI 1066
 DB 1002 nsmrrggaatphtpaaavtigrvvidsapslphllllhmqrastvhllgdpqipai 1061
 QY 1067 DEEHAGLPAIRPELAPRSMMXVTHRCPADVCELRGAYPKIOTTSRYLRSFNNPEAIG 1126
 DB 1062 deehaglypaairpelaprsmmxvthrcpadvcelrgaypkioettsrylrsfnnpeaig 1121
 QY 1127 QKLVYTOAKAANPGALITVHEAGATFETTTIATADARGLIOSSRAHAYALTRHTEKC 1186
 DB 1122 qklyftgaakaanpgalstvheagatfettliatadargliqssrahaivaltrhtkc 1181
 QY 1187 VILDAPGLLRREVGISDVIVNNPFLAGSEVGHRSVTPRGMPDONLGTLOAPPSCOISA 1246
 DB 1182 vildapglrrrevgisdvivnnpflagsevghrsvtprgmpdonlgtloappscois 1241
 QY 1247 YHOLAEEIGHRRPAPVAAVLPCCPELEOGLLYMPOELTVSDSVLVEFELTDIYHCRMASQ 1306
 DB 1242 fhqlaeelighrrpapvaavlpccpeleqlllylpqelctcdsvvtfeltdivhcrmasq 1301
 QY 1307 RKAVLSTLVGRYGRRTKLYEAHSDVRESLARFPTTPPVQATTCELYELVANEVKGOD 1366
 DB 1302 rkavlstlvgryrtrtklyeahsdvreslarfpptppvqattcelyelvanevkgo 1361
 QY 1367 GSAVLELDLCNRDYSRTFPOKXCNKFTTGTEIHAHGKVGOGISAMSKFCALFGWPFPAI 1426
 DB 1362 gsavleldlcsrdysrtfpoxcnkfttgteihahgkvvgogisamskfcalfgwptfai 1421
 QY 1427 EKEITALLPNIIFYGDAYEESVFAAASVAGSAGSCWVFNDFSEPDSTONFSLGLECVME 1486
 DB 1422 ekaitallpniifygdavtfdvfaaasvagsagscwvfndfsepdstgnfslglecalme 1481
 QY 1487 ECGMPQHLIRLXHLVRSWMLIQARKESLKGFWKXSGSPGTLNNTVNMALITAHCFER 1546
 DB 1482 ecgmpqhlirlyhlrseswmlqarkeslrgfwkxsgspgtlntvnmavlitahcydfr 1541
 QY 1547 DFRVAEFKDDSVYLCSDYROSRNAALIIACGGLKLKVDYPRIGLYAGVVAAPGIGTLPD 1606
 DB 1542 dfrvaefkddsvylcsdyrosrnaaliiacgglklkvdypriglyagvvaapigalt 1601
 QY 1607 VYRFAGRLSEKNMGPGPEPABOLRLAVCDLRLGLTNVAQVCVDVVSRYGVSPGLVHNL 1666
 DB 1602 vyrfagrlseknmgpgpepaeqlrlavsdflrltnvaqmcvdrvsvrygvsplvnhli 1661
 QY 1667 GMLDTIADGKAHFTETIKPVLDLTNSIIORVE 1698
 DB 1662 gmldavdgkahnftesvkvpidltlnsilicrve 1693

RESULT 5
 W80196
 ID W80196 standard; Protein; 1693 AA.
 XX
 AC W80196;
 XX
 DT 23-DEC-1998 (first entry)
 XX
 DE Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.
 XX
 KW Enterically transmitted non A non B hepatitis virus; ET-NANB;
 KW Hepatitis E virus; HEV; Burma HEV isolate; vaccine;
 XX diagnostic probe.
 XX

OS Non A non B Hepatitis virus.
 XX US5824649-A.
 PN 20-OCT-1998.
 XX 07-JUN-1995; 9505-0475807.
 PF 25-JUL-1994; 9405-0279823.
 XX 17-JUN-1988; 8805-0208997.
 PR 11-APR-1989; 8905-0336672.
 PR 16-JUN-1989; 8905-0367486.
 PR 13-OCT-1989; 8905-0420921.
 PR 05-JUL-1990; 9005-0505888.
 PR 07-JUN-1995; 9505-0475807.
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 PA Bradley DM, Fry KE, Krawczynski KZ, Reyes GR, Tam A;
 PI Yarbough PO.
 XX WPI: 1998-582599/49.
 DR N-PSDB: V66321.
 PS Hepatitis E virus proteins - useful for diagnosis or vaccine
 PT production the virus
 XX Claim 22: Columns 57-66; 47pp; English.
 CC W80196-98 are encoded by the genome of the Burma strain of
 CC enterically transmitted non A non B hepatitis virus (ET-NANB)
 CC (hepatitis E virus (HEV)). The specification describes an isolated
 CC protein which is specifically immunoreactive with antibodies present
 CC in individuals infected with HEV and encoded by a sequence contained
 CC in an open reading frame (ORF) of an HEV genome. The genome has a
 CC sequence that is more than 70% identical to the ORF1 sequence from
 CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
 CC probe for ET-NANB.
 XX Sequence 1693 AA:
 SQ

Query Match 82.6%; Score 7448; DB 19; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 1 PGTTTAAEQALAAANALANAVVVPFLSRVOTELINIMOPROLVFEPEVLMHPRIOR 60
 DB 10 pglttatagaalaanaalanavvvpflshqgfellinimprqlvfrpevfmbpigr 69
 QY 61 VHNLEEQYCRARARCLEYGAHPRISINDPNVLRHCFELRPVGDVORWYSAPTRGPAAN 120
 DB 70 vhnleelycrarsrcleigahprisindhpnvhrclfrpvgdvgrwylaptrgpaan 129
 QY 121 CRSSALRGLEPPADRTYCFDFSRCAFAEEGVALYSHDLMPADVAAAMARHGTRIXAA 180
 DB 130 crssalrtglpadrtycldfsgcnpaetgialyslhdmspsdvaeamtrhgmtrilxaa 189
 QY 181 LHTPEPEVLPPGTVHTSYLLIHGDRAVVTYEGDTSAGYNHDSILRAMIRTKTIYGDH 240
 DB 190 lhtppevlppgtyhtsyllihdgravvttyegdtsagynhdslrmtlrrtktygdh 249
 QY 241 PLVIERVRAIGCHRVLLTLTAPEBSPMPVYVPRSTFEVYRSIFPGGSPSLPEPASCSTK 300
 DB 250 plvieriraisgchrvllltaapebpmvypvprstfevyrsifpggspslpescstk 309
 QY 301 STEHNAVHVMIDRLMLGATLDOAFCCSRMTYRISYKVMYGVALVAMGNASDAL 360
 DB 310 stehnavahvmdrlmlgatlldoafccsrmtlyriskyvmtygvallvamegnasdal 369
 QY 361 TAAITAAVLTICRQRYLRTQASIKGMRLGVENAKQKITRILYSWLFEPKSGRDYIPGROLQ 420
 DB 370 tavitaaylvlticrqrylrtqaisikgmrlgvnenaqkkitrilyswlfepeksgrdyipgrolq 429

QY 421 FVAQCRRLWSAGPHLDPRVLYPEDESVPCCRRTLTAKVAGKFCQCFMRMLGQECTCELEPAE 480
 DB 430 fvaqcrwlwsagphldprvlypeidesvpcrcrtrltakvagskfcqcfmrmlgqectcfelepa 489
 QY 481 GLVGDHGHNEAEVSGSEVPAPBAHLDVSGTVVHGHOLEALYRALNVPODIAARASRLT 540
 DB 490 glvgdghghneaevgsevpapbaahldvsgtvtvhgholealyrainvpoodiaarasrlt 549
 QY 541 ATVELVASPDRLCEKRTVLGNKPTFTVVDGAHLNANGPEQVYLSFPDASRQSMGASHSLT 600
 DB 550 atvelvaspdrlcekrvtlgnkptftvvdgaahlngpeqvylsfpdasrqsmgashslt 609
 QY 601 YELTPAGLOYRISNGLDCTATPPEGAPSAAGVEVAAPFASALYRNRRTQRSLTGIM 660
 DB 610 yeltpagloyrisngldctatppegapasaagvevaapfalsalyrnrtrqrslltgim 669
 QY 661 LHPEGILGIFPPSPGIMESANPFCGECTLYRTWS-TSGFSSDSP-----PEAAPA 714
 DB 670 lhpegilgifppspgimesanpfcgectlyrtwstsgfssdsp-----peaapa 719
 QY 715 MAATPGLPHSTPPSDIWLPLPSEEFVDAAEV-PPAPDPAGLP-GPVVLTPPPPPVYH 772
 DB 730 raatppl-----aatpppdpappspapalaepasgatag 766
 QY 773 KPSIP-PPSRNRRLTYTPDGAAVYAGSLFESPCDMLVNASNGHRPGGGLCHAFYORP 831
 DB 767 apatthqtatrhrltlftypgskvfaslfesctclvnaasnvdhprpgglchafyoryp 836
 QY 832 EAFYPTFEIMRECLAAVTLPRPIHAHAVADYVEONPKLEAAVRETCRRGTAAVPLT 891
 DB 827 asfdaasfmrtdgaagatclprplihavadytlempkrlaayretcsrlytaapll 886
 QY 892 GSGIYQVPSLSDAMERNRHPDELYLTPEANWEANRPAOPVLTITEDTARTANAL 951
 DB 887 gsgiyqvpisdsdamerhnrhpdelyltpeanweanrpaopvltittdartanal 946
 QY 952 EIDATEVGRACAGCTISPEIVHYQFTAGVPSGSKSISIOGSDVVDVYVTRRLNSMR 1011
 DB 947 eldsatvgracagctrvtpgvvyqftagvpsgsksrslsqadavdvvyvtrrlnawtr 1006
 QY 1012 RGFPAFTPHTAARTYIGRRVVIDEAPSLPHLLLHMQRASVYHLGDPNQIPADIFEHA 1071
 DB 1007 rgfaafthptaartytgrrvvideapslphlllmqraatvhllygdpnqipadifeha 1066
 QY 1072 GLVPARPELAPTSMAWVTHRCPADVCELIRGAYPKIOTTSRYLSLFENNEPAIGOKLYX 1131
 DB 1067 glvpairpdlgpcswvthrwpadvcclirgaypmqttsrlyrsifwepavagqlvf 1126
 QY 1132 TOAKAANPCATIVYHEAQATFTETITATADARGLIOSSRAHAYALVTHTEKCYTLDA 1191
 DB 1127 tgaakpanpsvsvcheagatytetltatadargllqssrahaivalthtckcvtlida 1186
 QY 1192 PGLIREVIGISDVIVNNEFLAGGEVGHXRPVIRGNPDOLGTLQAPPSCOISAVYHQLA 1251
 DB 1187 pglirevigisdvivnnflaggevgxhrrpvirgnpdolgtlaqppscoisavyhqla 1246
 QY 1252 EELGHRPAPVAAVLPCEPEOGILLYMPQELTVSDSVLVELNDIVHCRMAAPSORKAVL 1311
 DB 1247 eelghrpvpaavlpcepeogillympqeltdsvsvlvelndivhcrmaapsqrav 1306
 QY 1312 STLVGKRGRTKYLEAHSVRESLARFTPTIGPVQATTEDELYELVAMWKEGQDGSAYL 1371
 DB 1307 stlvgrgrrtkyleahsvreslarfptigpvqattedelylvamwkeqdgdsayl 1366
 QY 1372 ELIDCNRDVSRIPEFOXCKKFTTGERTIANGKVGOGISANSKFFCALFGMPRAIEKEIL 1431
 DB 1367 elidcnrdsrileffqdcckfttgetiangkvvggaisaskffcalfgpwaiaekail 1426
 QY 1432 ALLPNIYGDAYEESVFAAASVAGSAGCWVFNDFSEFSDTONNFSGLGECVYMEECGMP 1491
 DB 1427 allpnigydayeessvfaaasvagsagcwvfnfsefstdtgnfsglgecalmeecgmp 1486

QY 1492 CMLILYHLVRSAMILQAPKESLKGFWKHSCEPGLLMNTVMNMAIIACHYEFDERVA 1551
 DB 1487 GMLILYHLVRSAMILQAPKESLKGFWKHSCEPGLLMNTVMNMAIIACHYEFDERVA 1546
 QY 1552 AFKGGDSVLCSDYRQSRNAALINGCGLKLVDRPILGLYGVVAVPGLTLPVVRPA 1611
 DB 1547 afkgdssvlycsgdyrqpaaavilaggcglklyvdfrlglyagvavapglpavifa 1606
 QY 1612 GRLSKKNMGCPGERAEQLRLAVCDLRLGTLTNVAQVCVDVSVFVSPGLVHNLIGMLQT 1671
 DB 1607 glltekngvpgpereaegjrlavsdflrklitvaqmcvdsrvyvspglvnhlignlmda 1666
 QY 1672 IADGKAHFEETIKPVLDTLNSTIORVE 1698
 DB 1667 vadgkantesvkvpyldltnsilcrve 1693
 RESULT 6
 B24119
 ID B24119 standard; Protein; 1693 AA.
 XX B24119;
 XX 29-JAN-2001 (first entry)
 DE Hepatitis E virus Burma strain ORF1 protein sequence SEQ ID NO:7.
 XX
 KW Hepatitis E virus; HEV; non-A/non-B viral hepatitis; viral protein;
 KM immunological; diagnosis; hepatitis; infection; identification;
 KM detection; immunoreactive; hepatotropic; antiinflammatory; virucide;
 XX vaccine; antiviral; antigenic; antibody; antigen.
 OS Hepatitis E virus.
 XX
 PN US6120988-A.
 PD 19-SEP-2000.
 XX
 PF 07-JUN-1995; 950S-0478507.
 XX
 PR 25-JUL-1994; 94US-0279823.
 PR 05-APR-1991; 91US-0681078.
 PR 17-JUN-1988; 88US-0208997.
 PR 11-APR-1989; 89US-0336672.
 PR 16-JUN-1989; 89US-0367486.
 PR 13-OCT-1989; 89US-0420921.
 PR 05-APR-1990; 90US-0505888.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PA (USGO) US GOVERNMENT.
 XX
 PI Yarbough PO, Krawczynski KZ, Fry KE, Bradley DW, Tam A, Reyes GR;
 DR N-PSDB; A55259.
 XX
 PT MPI: 2000-593712/56.
 PT
 PT Identifying recombinant antigen immunoreactive with antibody induced by
 PT hepatitis E virus (HEV), for detecting HEV infection, comprises
 PT immunoreacting a polypeptide from an HEV genome with an HEV-positive
 PT antiserum -
 XX
 PS
 PS Claim 6; Column 55-64; 46pp; English.
 XX
 CC The present invention describes a method for identifying a recombinant
 CC antigen immunoreactive with a hepatitis E virus (HEV)-induced antibody.
 CC The method comprises producing a polypeptide derived from an HEV genome,
 CC immunoreacting the polypeptide with an HEV-positive antiserum and
 CC selecting the polypeptide as a recombinant antigen if the polypeptide
 CC reacts with the HEV-positive antiserum. The method is useful for
 CC identifying recombinant antigen immunoreactive with antibody induced by
 CC HEV. The enterically transmitted non-A/non-B hepatitis (ET-NANB (also
 CC known as HEV))-specific fragments are useful for identifying ET-NANB-
 CC derived cDNAs, which contain additional sequence information, as primers

CC for detecting ET-NANB viral genomic material in a patient sample, for
 CC the synthesis of polypeptides for use in immunoassays, and for
 CC identifying similar antigenic regions encoded by related viral strains,
 CC e.g. Burmese strain. The antigens are especially useful in the
 CC preparation of vaccine against ET-NANB infection. These antigens may
 CC further be used to prepare antibodies to ET-NANB virus particles for
 CC use directly as antiviral agents, and to produce antiserum designed for
 CC pre- or post-exposure prophylaxis. The present sequence represents a
 CC specifically claimed HEV Burma strain protein sequence for use in the
 CC present invention.
 CC
 XX
 SO Sequence 1693 AA:
 Query Match 82.6%; Score 7448; DB 21; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;
 QY 1 PGITTAIOAALAAANSALANAVVVRPPLSRVQTEILINMQPQOLVFRPEVLNHP1QR 60
 DB 10 pgitatagaalaansalanavvrplshqglelllnmqprqlvfrpevfwnhp1qf 69
 QY 61 VIHNELEQYCRARAGRCLEVGAPHSINDNPNVLIHRCFLRPGVGRDQVQWYASAPTRGPAAN 120
 DB 70 vihnelelcrarsgrclelghaprsindpnvvrcllrpgvgrdqvgrwyacptrgpaan 129
 QY 121 CRRSALRGLPPADRTYCFDFGFSRCAFAETGVALYSLDHPADVAEAMRHGXTRLYAA 180
 DB 130 crrsalrglpadrtlycldfgsgcnfaetgialyslhdmpsdvaeamrfhgmtrlyaa 189
 QY 181 LHLPEVLLPRTYHTTYSYLLIHGDRAVVTYEGDTSAGVNHDSILRAWIRTKIYGDH 240
 DB 190 lhlpevllprtyhtyasyllihgdrravvtyegdtsagvnhdsvnlrswirtklygdh 249
 QY 241 PLVIERVAICGFVLLITAPRPSMPYVPRSTEVYVNSITGPGSPLFSASCTK 300
 DB 250 plviervalcghvllitaprpssmpyvrstevyvrslitgpgspflfsactk 309
 QY 301 STFAVAVPHIMDRMLFGATLDDOAFCCSRMTYLRGISYKVYTGALVANAGMNASDAL 360
 DB 310 stfnavavphidrmldfgatlddgaefccsrmtlylrghsykvvtgvlvanegmnsadal 369
 QY 361 TAXITTAAYLTCHORYLQTAISKGMRLGVEHAKFTRLYSWLEFGSGHDYIPGROLQ 420
 DB 370 taxitaaayltchorylqtaiskgmrlgvehakfttrlyswlfeqsgdyipgrql 429
 QY 421 FYACCRMLSGFHLDPRLVLFEDSVPCRCRFTPKKVAAGKCCMRMLGOCCTELPAE 480
 DB 430 fyaccrmlsgfhlprlvlfedsvpcrcrftpkkvaagkccmrmlgocctcltpae 489
 QY 481 GLVGDHGDNEAYGSEVDPAPRALDYSGTYAVHGOLEALYRALVNPDIARASRLT 540
 DB 490 glvgdghdneaygsedvparsaldysgtavhgolesalvnpdiaravagrlt 549
 QY 541 ATVELVASPDRLERCTYLGNKTFRTTYVDCAHLEANGDEQVLSFSDASROSMGAGSHSLT 600
 DB 550 atvsvsqvdrldcettlgnktrftrtyvdcavhlangdeqvlsvfssdasrsmagpshlt 609
 QY 601 YELTPAGLOVRISNGDCATPPPGGAPSAPEVNAFCALRYRNPQRHSLTGLGW 660
 DB 610 yaasaglevryvaagldhnavtagvpsrpsapevnafcallyrlnaqrhslignlw 669
 QY 661 LHPGGLGIPPEPSPGHIMESANPCEGLTYRTWS-TSGFSSDPSF-----PEAAPA 714
 DB 670 fhpgglgltpfsgphimwseanpfcgsltyrttwseavaspapdlgfmsepsips 729
 QY 715 MAATPGLPBSTPYSDIWLVRPPESEFQVDAAY-PAAPPAALP-GRVVLTPPPPPVH 772
 DB 730 raatppl-----aaipppappppspsapalaepasgtag 766
 QY 773 KPSPR-PPSNRRLLLYRPGAKVYAGSLPESODMWLVNNSPGRGGGLCAHAFYGRF 831
 DB 767 apaltbqatnrlitlcypgskvflagsltesctcwlvnasnvdrp99lchaflgqfyp 826


```

Db 250 p|v|e|r|a|t|g|c|h|t|v|l|l|t|a|e|p|s|p|m|y|v|y|p|r|e|v|y|s|i|f|g|p|g|t|p|l|f|t|s|c|s|k 309
QY 301 S|T|P|A|V|P|H|I|D|R|I|M|L|F|G|A|T|D|D|A|F|C|S|R|L|A|T|L|R|G|S|Y|V|V|T|G|A|V|A|N|G|A|N|S|E|D|A|L 360
Db -310 s|t|r|i|a|v|p|h|i|d|r|i|m|l|f|g|a|t|d|d|a|f|c|s|i|m|y|l|g|s|y|v|v|t|g|a|v|a|n|g|a|n|s|e|d|a|l 369
QY 361 T|A|I|T|A|V|L|T|C|H|O|R|Y|L|N|T|Q|A|I|S|K|M|R|L|G|V|E|H|A|O|K|E|T|R|L|S|W|L|F|E|K|S|G|N|D|Y|I|P|G|R|O|L 420
Db 370 t|a|i|t|a|v|l|t|c|h|o|r|y|l|n|t|q|a|i|s|k|m|r|l|g|v|e|h|a|o|k|e|t|r|l|s|w|l|f|e|k|s|g|n|d|y|i|p|g|r|o|l 429
QY 421 F|Y|A|O|C|R|M|L|S|G|F|H|D|R|V|L|F|D|S|V|P|C|R|T|F|L|K|K|Y|A|G|K|C|C|M|R|L|G|O|E|C|T|F|L|E|P|A|E 480
Db 430 f|y|a|o|c|r|m|l|s|g|f|h|d|r|v|l|f|d|s|v|p|c|r|t|f|l|k|k|y|a|g|k|c|c|m|r|l|g|o|e|c|t|f|l|e|p|a|e 489
QY 481 G|L|V|D|H|G|H|D|N|E|A|Y|G|S|E|V|D|P|E|A|H|L|V|S|G|T|V|A|V|H|Q|L|E|Y|A|L|V|P|O|D|I|A|R|A|S|R|L|T 540
Db 490 g|l|v|d|h|g|h|d|n|e|a|y|g|s|e|v|d|p|e|a|h|l|v|s|g|t|v|a|v|h|q|l|e|y|a|l|v|p|o|d|i|a|r|a|s|r|l|t 549
QY 541 A|T|V|E|L|V|A|S|P|R|L|E|C|R|T|V|L|G|N|K|T|F|T|V|D|G|A|H|L|E|A|N|G|P|E|Q|V|L|S|F|D|A|S|R|O|S|G|A|S|H|L|T 600
Db 550 a|t|v|e|l|v|a|s|p|r|l|e|c|r|t|v|l|g|n|k|t|f|t|v|d|g|a|h|l|e|a|n|g|p|e|q|v|l|s|f|d|a|s|r|o|s|g|a|s|h|l|t 609
QY 601 Y|E|L|T|P|A|L|O|R|I|S|N|G|D|C|T|A|T|P|P|G|A|P|S|A|P|G|E|V|A|A|F|G|A|L|R|Y|R|N|F|T|O|R|H|L|T|G|L|W 660
Db 610 y|e|l|t|p|a|l|o|r|i|s|n|g|d|c|t|a|t|p|p|g|a|p|s|a|p|g|e|v|a|a|f|g|a|l|r|y|r|n|f|t|o|r|h|l|t|g|l|w 669
QY 661 L|H|P|G|L|G|I|P|P|P|S|P|G|H|I|W|E|S|A|N|P|C|E|G|T|Y|R|T|W|S|-----T|S|G|S|D|F|S|P|E 709
Db 670 l|h|p|g|l|g|i|p|p|p|s|p|g|h|i|w|e|s|a|n|p|c|e|g|t|y|r|t|w|s|-----t|s|g|s|d|f|s|p|e 724
QY 710 A|A|P|R|A|A|A|T|E|G|L|P|H|S|T|P|P|S|D|I|W|L|P|P|S|E|E|Q|V|D|A|A|P|-P|R|A|D|P|A|-G|L|P|V|L|T|P|P 767
Db 725 a|a|p|r|a|a|t|e|g|l|p|h|s|t|p|p|s|d|i|w|l|p|p|s|e|e|q|v|d|a|a|p|-p|r|a|d|p|a|-g|l|p|v|l|t|p|p 761
QY 768 P|P|R|V|H|R|S|I|P|-P|S|R|N|R|R|L|Y|T|Y|P|D|G|A|K|Y|V|A|G|S|L|F|E|S|D|C|M|L|V|N|A|S|N|P|R|H|R|G|G|L|H|A|P 826
Db 762 p|p|r|v|h|r|s|i|p|-p|s|r|n|r|r|l|y|t|y|p|d|g|a|k|y|v|a|g|s|l|f|e|s|d|c|m|l|v|n|a|s|n|p|r|h|r|g|g|l|h|a|p 821
QY 827 Y|O|R|E|P|E|A|F|T|E|P|E|I|R|E|G|L|A|Y|L|T|P|R|P|I|H|A|V|A|D|Y|R|V|E|O|N|P|R|L|E|A|A|Y|R|E|T|S|R|R|G|T|A 886
Db 822 y|o|r|e|p|e|a|f|t|e|p|e|i|r|e|g|l|a|y|l|t|p|r|p|i|h|a|v|a|d|y|r|v|e|o|n|p|r|l|e|a|a|y|r|e|t|s|r|r|g|t|a 881
QY 887 A|Y|P|L|G|S|G|I|Q|V|P|V|S|L|F|D|A|W|E|R|N|H|R|P|G|D|E|L|Y|T|E|P|A|N|F|E|A|K|P|A|Q|V|L|T|T|E|D|R|A|T 946
Db 882 a|y|p|l|g|s|g|i|q|v|p|v|s|l|f|d|a|w|e|r|n|h|r|p|g|d|e|l|y|t|e|p|a|n|f|e|a|k|p|a|q|v|l|t|t|e|d|r|a|t 941
QY 947 A|N|L|E|I|D|A|T|E|V|R|A|C|G|C|T|I|S|P|G|I|V|H|O|F|T|A|G|V|P|G|S|R|S|I|O|G|D|V|V|V|V|P|R|E|R|L 1006
Db 942 a|n|l|e|i|d|a|t|e|v|r|a|c|g|c|t|i|s|p|g|i|v|h|o|f|t|a|g|v|p|g|s|r|s|i|o|g|d|v|v|v|v|p|r|e|r|l 1001
QY 1007 N|S|W|R|R|G|F|A|F|T|P|R|A|V|T|I|G|R|R|V|I|D|E|A|P|S|L|P|R|H|L|L|H|M|R|A|S|S|V|H|L|G|D|P|N|Q|I|P|A|I 1066
Db 1002 n|s|w|r|r|g|f|a|f|t|p|r|a|v|t|i|g|r|r|v|i|d|e|a|p|s|l|p|r|h|l|l|h|m|r|a|s|s|v|h|l|g|d|p|n|q|i|p|a|i 1061
QY 1067 D|E|H|A|G|I|V|P|R|E|P|A|T|S|W|X|Y|T|H|R|C|P|A|D|V|C|E|L|R|A|Y|K|I|O|T|T|S|V|L|S|L|E|N|E|P|A|I|G 1126
Db 1062 d|e|h|a|g|i|v|p|r|e|p|a|t|s|w|x|y|t|h|r|c|p|a|d|v|c|e|l|r|a|y|k|i|o|t|t|s|v|l|s|l|e|n|e|p|a|i|g 1121
QY 1127 O|K|L|V|T|Q|A|A|A|N|G|A|T|V|H|E|A|O|A|F|T|E|T|T|A|T|A|R|G|I|O|S|R|H|A|I|V|A|L|T|R|H|E|K 1186
Db 1182 o|k|l|v|t|q|a|a|n|g|a|t|v|h|e|a|o|a|f|t|e|t|t|a|t|a|r|g|i|o|s|r|h|a|i|v|a|l|t|r|h|e|k 1181
QY 1187 V|I|D|A|R|G|L|E|V|G|I|S|D|V|I|N|N|F|L|A|G|E|V|G|X|H|R|S|V|I|R|P|R|N|P|O|N|L|G|T|L|O|A|F|P|R|S|C|O|I|S|A 1246
Db 1182 v|i|d|a|r|g|l|e|v|g|i|s|d|v|i|n|n|f|l|a|g|e|v|g|x|h|r|s|v|i|r|p|r|n|p|o|n|l|g|t|l|o|a|f|p|r|s|c|o|i|s|a 1241
QY 1247 Y|H|O|L|A|E|R|G|R|R|P|A|V|A|V|P|R|C|E|L|E|G|L|X|M|Q|E|L|T|V|S|S|V|L|F|E|L|T|D|I|V|H|R|C|H|A|P|S|Q 1306
Db 1242 y|h|o|l|a|e|r|g|r|r|p|a|v|a|v|p|r|c|e|l|e|g|l|x|m|q|e|l|t|v|s|s|v|l|f|e|l|t|d|i|v|h|r|c|h|a|p|s|q 1301
QY 1307 R|K|A|V|L|S|T|V|R|G|R|R|T|L|Y|A|H|S|D|V|R|E|S|L|R|F|I|P|T|G|Q|A|O|A|T|C|E|L|Y|E|A|M|V|E|A|N|E|G|O|D 1366

```

```

Db 1302 r|k|a|v|l|s|t|v|r|g|r|r|t|l|y|n|a|s|h|d|v|r|d|s|l|r|f|i|p|i|a|g|p|v|t|t|c|e|l|y|e|e|a|m|w|e|k|g|d 1361
QY 1367 G|S|A|V|E|L|D|L|C|R|N|D|V|R|T|F|P|O|K|X|N|K|F|T|G|E|T|A|H|G|V|G|O|I|S|A|W|S|K|T|F|C|A|L|P|G|P|R|A|I 1426
Db 1362 g|s|a|v|e|l|d|c|r|n|d|v|r|t|f|p|o|k|x|n|k|f|t|g|e|t|a|h|g|v|g|o|i|s|a|w|s|k|t|f|c|a|l|p|g|p|r|a|i 1421
QY 1427 E|K|E|I|L|A|L|P|N|I|F|Y|G|A|Y|E|E|S|V|F|A|A|V|S|G|A|S|C|W|F|E|N|D|S|E|P|D|S|T|O|N|N|F|S|L|G|E|C|Y|V|E 1486
Db 1422 e|k|e|i|l|a|l|p|n|i|f|y|g|a|y|e|e|s|v|f|a|a|v|s|g|a|s|c|w|f|e|n|d|s|e|p|d|s|t|o|n|n|f|s|l|g|e|c|y|v|e 1481
QY 1487 E|C|G|P|O|L|R|L|H|L|V|R|S|A|W|L|O|A|P|K|E|S|L|K|G|F|W|K|H|S|E|P|G|T|L|L|N|Y|W|N|A|I|I|A|H|C|Y|E|R 1546
Db 1482 e|c|g|p|o|l|r|l|h|l|v|r|s|a|w|l|o|a|p|k|e|s|l|k|g|f|w|k|h|s|e|p|g|t|l|l|n|y|w|n|a|i|i|a|h|c|y|e|r 1541
QY 1547 D|F|R|V|A|F|K|G|D|S|V|V|C|D|Y|R|O|S|R|N|A|A|L|I|N|G|C|G|L|K|L|V|D|R|P|G|L|Y|G|V|V|A|G|L|T|R|P|D 1606
Db 1542 d|f|r|v|a|f|k|g|d|s|v|v|c|d|y|r|o|s|r|n|a|a|l|i|n|g|c|g|l|k|l|v|d|r|p|g|l|y|g|v|v|a|g|l|t|r|p|d 1601
QY 1607 V|R|P|R|A|G|R|L|S|E|K|N|G|P|G|R|A|E|Q|L|R|A|V|C|D|F|L|R|G|L|T|N|V|A|Q|V|C|D|V|S|R|V|Y|S|P|G|L|V|H|N|L| 1666
Db 1602 v|r|p|r|a|g|r|l|s|e|k|n|g|p|g|r|a|e|q|l|r|a|v|c|d|f|l|r|g|l|t|n|v|a|q|v|c|d|v|s|r|v|y|s|p|g|l|v|h|n|l 1661
QY 1667 G|M|Q|T|A|D|G|K|A|H|F|E|R|T|K|P|V|L|D|I|N|S|I|O|R|V|E 1698
Db 1662 g|m|q|t|a|d|g|k|a|h|f|e|r|t|k|p|v|l|d|i|n|s|i|o|r|v|e 1693

```

```

RESULT 8
W76368
ID W76368 standard: Protein; 1693 AA.
AC W76368;
XX
DT 03-DEC-1998 (first entry)
DE Hepatitis E virus hollow particle protein #1.
KW Hollow particle protein; virus; antibody; detection; immunoassay;
KW Infection.
OS Hepatitis virus.
XX
FH Key Location/Qualifiers
FT Protein 1..1693
FT /note="Partial sequence"
PN JP10234383-A.
PD 08-SEP-1998.
PE 28-FEB-1997; 97JP-0062445.
PF 28-FEB-1997; 97JP-0062445.
PR 28-FEB-1997; 97JP-0062445.
XX
PA (DENK-) DENKA SEIKEN KK.
PA (KOKU-) KOKURITSU YORO ETSEI KENKYUSHO.
DR N-PSDB; V61687.
XX
PT Hepatitis E virus hollow particle poly(peptide(s) and nucleic acids
PT encoding it - useful for more accurate detection of HEV in samples,
PT using immuno-assays and nucleic acid hybridisation
XX
XX Claim 10: Page 17-24; 29pp; Japanese.
XX
XX This sequence represents a Hepatitis E viral hollow particle protein.
XX This polypeptides can be used to raise antibodies to detect HEV
XX infection in samples, e.g. by immuno-assay based techniques, and the
XX nucleic acid can be used for the same in nucleic acid hybridisation
XX assays. The polypeptides and nucleic acids allow more accurate
XX detection of HEV than previously possible.
XX

```



```

XX  US5789559-A.
PN
XX
PD  04-AUG-1998.
XX
PF  25-JUN-1994; 94US-0279823.
XX
PR  05-APR-1991; 91US-0681078.
PR  17-JUN-1988; 88US-0208997.
PR  11-APR-1989; 89US-0336672.
PR  19-JUN-1989; 89US-0367486.
PR  13-OCT-1989; 89US-0420921.
PR  05-APR-1990; 90US-0505888.
PR  25-JUL-1994; 94US-0279823.
XX
PA  (GENE-) GENELABS TECHNOLOGIES INC.
PI  Bradley DM, Fry KE, Krawczynski KZ, Reyes GR, Tam A;
PI  Yardough PO;
XX
DR  WPI: 1998-446186/38.
DR  N-PSDB: V54729.
XX
PT  Hepatitis E virus DNA - useful for e.g. virus detection and viral
PT  protein production
PS  Disclosure: Columns 55-64; 45pp; English.
XX
XX  W71209-11 represent the proteins encoded by the open reading frames
CC  (ORFs) of the DNA sequence of the Burmese isolate of an enterically
CC  transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic
CC  acid sequence may be used for identifying and sequencing the entire
CC  viral agent (also referred to as HEV), detecting ET-NANB in
CC  infected samples, e.g. by specific amplification of virus-derived DNA
CC  sequences and for producing recombinant viral proteins for use in
CC  vaccines.
XX
SQ  Sequence 1693 AA;

```

```

DB  430 fyaqgrtwlsagfhnldpvlvfdesapchortairkalsklocfmkwyggetctclqpae 489
OY  481 GLVGHGHDNENAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNPDIARASRLT 540
DB  490 gavgqgdnneayegsdvdpesaalsdsyvvpgtaldpilyqaldipaelvaragrilt 549
OY  541 ATVELVASPDRLRECTVVGKNTFRTTVDGALHLENANGSEOVLSFSDASROSGAGSHSLT 600
DB  550 atkvsgvqgdrlidcettlgnkftitsfydgavlenpserlmisfdaasctmaagpfslt 609
OY  601 YELPPAGLQVRISSNGLDCTATFPFGAPSAAPGEVAACFSCALYRYNRTGHSITGSLG 660
DB  610 yaasaaglevryaaagldhravfapgyvrsapgevtalcasyfyrineaqrhslignlw 669
OY  661 LHPEGLLIFPPFSGHGHWESANPRCGCTLYTRKWS-TSGFSSDFSP-----PEAAPA 714
DB  670 fhpegllglftapfsgphwesanpfcgectlytrtwsevdavssparpdlgfmsepslps 729
OY  715 MAATPGLPHSTPPVSDIWLPPPSSEFQVDAAPV-PPAPDPAGLP-GVVLTPPPPEVH 772
DB  730 raaptl-----aa1pppapdpsspapalaepasgatag 766
OY  773 KPSIP-PPSRNRRLTYTPDGAKYAGSLFESDCDMLVNASNPGHRRPGGLCHAFYORFP 831
DB  767 apaltghqarhrillftypdgskvfagslftectclwnasnvdhprpggylchafyqtyp 826
OY  832 EAFYPTFIMEEGLAAYVLTLPRIIHAVAPDYRVQNFKLEAAVRETCSRRGTAAVPL 891
DB  837 asfdaasfvmddgaagaelcprlphavapdyrlhmpkrlleaayretcsrlgaayp11 886
OY  892 GSGIYQVVSLSFDAMENHNRPGDELTYTEPAAMWFEANKRAOVLTTTEGTARTANL 951
DB  887 gtglyqvpigsfdaewenhnpodelylpelaalaafveantrp1lttevatant1ai 946
OY  952 EIDATEVGRACAGCTSPGIVHYQFTAGVPGSGKSRSIQOGVDVYVVPTRDLNSMR 1011
DB  947 eldsatdvgracagcrvpyvvyqftagvpgsgksrsltgadvdvvyvpreltnawr 1006
OY  1012 RGFAPFPTHTAARTYTGRRVYIDAPSLPRHLLLLHMGASVHLHGDNPQIPIADEPHA 1071
DB  1007 rgaafthphtaavtqgrvvvdaaps1pnh111lmgraatvhl1gdqnp1paldelha 1066
OY  1072 GLVPAIRPELPATPMWXXVTHRCPADVCELRGAPYKIDOTTGRVRSLEFNPAAGOKLVX 1131
DB  1067 glvpa1rpd1pbtswvnhthrpadvcel1rgaypmq1tser1vrs1fwgpa1vq1kvf 1126
OY  1132 TQAAKANPGAITYVHEAGATFTETIIATADANGLIQSSRAHAIVALTTRHTEKCVILDA 1191
DB  1127 tqaakpangsvtvheaqalyletlliatadarg1lqssrahaiva1lrrhtekcvilda 1186
OY  1192 PGLIREVGISDVYNNPFLAGGEVGHRRPSYIPRGNPDONIGTIOAPPSCQISAIYHOLA 1251
DB  1187 pgl1revgisdaivnnfl1agge1ghrpsv1pgrpnpa1n1la1fppscq1sa1fng1a 1246
OY  1252 EELGHRPAPVAAVLPCELEEOGLYMPQELTVSDSVLPFLDTIVICRMAPAPQORAVL 1311
DB  1247 eelghrpyvaav1pce1eleog1ly1pge1ltcdsvv1fcltdiv1ncma1pqr1xav1 1306
OY  1312 STLVGRYGRRTKLYEAASDVRESIARFIPTIGVQATTCLEYELVAMVEKGGDGSAYL 1371
DB  1307 stlvgr1grtk1lyna1s1dv1res1ar1f1p1t1g1v1q1at1tcl1ey1el1vam1ve1k1gd1gs1ay1l 1366
OY  1372 ELIDCNRDVSRIITFFQKKCNKFTTGETIAGKVGQGISAMSKTFCALGFWFAIEKEIL 1431
DB  1367 elidc1n1dv1sr1it1ff1q1k1cn1k1ft1t1g1et1i1ag1k1vg1q1is1am1sk1t1fc1al1g1fw1fa1ie1ke1il 1426
OY  1432 ALLEPNIFYGDAVEEVSFAAASGASGCMVFENFSFSDSTONNFSGLECYVVAEEGMP 1491
DB  1427 allp1gv1fy1gda1dd1v1sa1va1a1kas1m1v1end1f1se1ld1st1qn1f1s1gl1e1oa1l1nee1c1gm1p 1486
OY  1492 QMLTRLHLVRSANVILQAPKESLGFPMKKHSGEGTLLMNTVMNAIITACYFERDRFVA 1551

```

Db 1487 gwllrllyhllrsawllgqpkesslrgfwxkhsgepglllwtvwmavithcydfdrfgva 1546
Qy 1552 AFKGDSDVVLCSDRQSNAAALLTAGCGLKLVYRPIGLYAGVYVAGLGTLPVYVFA 1611
Db 1547 afkgxksivlseyrqpagaavllagcgiklkvdfrpigliyagvvpajlgalpdrvifa 1606
Qy 1612 GRLSKNNPGPEREQLRLVCDPLRGLTNNVAVCVYVSVGVSGGLVHNLIGMLQT 1671
Db 1607 grltekwpqperaeqrlrlavsdllrlkltnvaqmcvavsrlygvspjlvhnlilgmlga 1666
Qy 1672 IADKAFHETIKPYLDLTNSIIOAVE 1698
Db 1667 vadgkafhtesvkvpldltnsilcrve 1693
RESULT 10
R51264
ID R51264 standard; Protein; 1693 AA.
XX
AC R51264;
XX
DT 21-OCT-1994 (first entry)
XX
DE HEV strain protein encoded by ORF-1.
XX
KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
XX antibody; detection; diagnosis; primates; stool suspension.
XX
OS Hepatitis E virus strain SAR-55.
XX
PN W09406913-A.
XX
PD 31-MAR-1994.
XX
PE 17-SEP-1993; 93WO-US08849.
XX
PR 18-SEP-1992; 92US-0947263.
XX
PA (USSH) US SEC DEPT HEALTH.
XX
PI Emerson SU, Purcell RH, Tsarev SA;
XX
DR WPI: 1994-118462/14.
XX
DR N-PSDB: Q45197.
XX
PT Purified hepatitis E strain SAR-55 virus - used to develop prods.
XX
PT for use in detection, diagnosis, vaccines and therapy of
XX
PT hepatitis E virus infection
XX
PS Disclosure; Page 57-62; 114pp; English.
XX
XX The sequences given in R51264-66 are encoded by the hepatitis E virus
XX (HEV) strain SAR-55. The CDNA sequence contains three open reading
XX frames (ORFs). These proteins can be used to stimulate the production
XX of protective antibodies upon injection into a mammal that would serve
XX to protect the mammal upon challenge with wild type HEV. The proteins
XX can be used for detection and diagnosis of HEV infection. The HEV
XX SAR-55 CDNA was isolated from primates inoculated with stool
XX suspensions obtained from hepatitis E patients.
XX
SQ Sequence 1693 AA:
Query Match 82.5%; Score 7436; DB 15; Length 1693;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 1396; Conservative 112; Mismatches 162; Indels 42; Gaps 7;
Qy 1 PCITATIAQAAALAAANSALANAVVVRPLSRVQTEILINMQPRQOLVRPEVLMMHPRIOR 60
Db 10 pgitataqaaalaanaaanaavvrplsrvgqlellnlmprijlvitpevfwhpbiqr 69
Qy 61 VHNLELEQYCRARAGRCLEVGAPHSINDNPVLRHRCFLRPVGRVQWRYSAPTGGPAAN 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 70 vihnleleycrarsgcrlcigahpsindnpvvrhrcflrpadrldvgrvtaptrgpaan 129
Qy 121 CRRSRLRLPLPADRTYCGDGSRCAPFAETGVALLYSLDMLPADYAEAMARHGXRLLAA 180
Db 130 crssrlrlplpadrtycldgtsrcapfaetgalyalsldmepsdvaeamfthgmkrlyaa 189
Qy 181 LHLPEVLLPQGTHTTSTYLLIHODGRAVVTYEGGTSAGYVHNDYSILAWIRTRTIVGDH 240
Db 190 lhlpevllppgytrtaaylllhodgravvtlyegdtlsagynhdvsnlswlrltkvqdh 249
Qy 241 PLVIERVRAICHFVLLTLTAAPESPMPYVPRSTEVYVSIPEGGSPSLFSPACSTK 300
Db 250 plviervraicghfvllltaapepmpyvprrstevyvisfpggtpslfprscstk 309
Qy 301 STEFAVPHVIMDRMLFEGATLDDQAFCCSRMLTYLRGISYKYTVGALVANSNMASDAL 360
Db 310 stfnavpahvdrmlfegatlddgafccsrmltylrgisykytvglvanegmasedal 369
Qy 361 TAXITRAVLTICHOHRYLRTQAIISKGMRLGVHAGKFTIRLYSMLFEKSGRDIYRGLO 420
Db 370 taxitraavltichqrylrltqalskgrmlrerenagkllrllyswlfeksgrdiyrglqe 429
Qy 421 FYAQCRMLISAGFHLDPVLVFEDESVPCCRTFLKRVAGKFCCEMRMLGOECTCLFEPAE 480
Db 430 fyaqcrmlisagfhlldprvlvfdesapchcrtalrkavskfccefmkvgldqectclfgpae 489
Qy 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEBALRYALNPQDIARASRLT 540
Db 490 glvgdghdneayegsdvdpaeasldisgyvypglalpyqaldipaeivaragrlt 549
Qy 541 ATVELVASPDRLRECTVYVGNTRFTYVDGAHLENGPEQYVLSFDASROSAGASHSLT 600
Db 550 atvvelvaspdrlrectvyntrfttyvdgaahleengpeqyvlsvfdasrosagashslt 609
Qy 550 atkvsvqvdgrldccelllgnkltfcsfydgavlecnperlmhlsdaasqstaagpifslt 609
Qy 601 YELTPAGVOVRISNGLDCTATFPFGAPSAAPGEVAFCSALRYRNFTQORHSLTGGLW 660
Db 610 yaasaaglevryaaagldhravfapgvprspgevtcfcsalryfnreaqlsltgfnw 669
Qy 661 LHPEGLIGIFPPFSPGHIMESANPPCGEGTLYTRTWS-----TSGFSDFSPPE 709
Db 670 fhpegllgfpafspghwesanpfcgesllytrtwsedavpapgdlgfts-----e 724
Qy 710 AAAPMAATPGLPHSTPPVSDIWLPPPSSEFQVNAAY-PPAPPA-GLRPVVLTPPP 767
Db 725 psipstraalp-----lp-----aaipppapdpdpstlspargpapp 761
Qy 768 PPPVHKPSIP-PPSRNRLLYTPDGAKYAGSLPESDCDMVYNASNPGRGGGLCHAF 826
Db 762 gatarapatlqtahrtrlllfypdgskvfagsltestctvlnasnvndhprggllchaf 821
Qy 827 YORPEAFYPTFEIMREGIAVYTLPRPIIHAADVREONPKRLDAVRETCSRGTGA 886
Db 887 yorpeafyptfeimregiaavtytlprpiihadvreonpkrlдавretcsrgrtga 881
Qy 887 AYPILGSGIYQVPLSLDPAWERNHRPGBELYLTPAANWFEANPQAPVYLTEDTART 946
Db 882 aypllgsgiyqvpplslpdawernhrpbdelyltpaawfeanpqpvcyltltedavart 941
Qy 947 ANLALEIDAATEVGRACAGCTISPGIVHYQTFAGYVPGSGSKRSIOGDVDDVVVYTRRLR 1006
Db 942 anlaleidsatdvgracagcrtvpgvvyqftagypsgskrstlqadddvvvtrrlr 1001
Qy 1007 NSMRRRGFAALTPTHTAARTYIGRRVYIDAEPSLPHLLILHOMASSVHLCDPNOIPAI 1066
Db 1002 nsmrrrgfaaltpthtaartygrrvyideapslphllilhmqravtllygdpnqipai 1061
Qy 1067 DEHAGIVPAIRPELAPISWVXVTHRCPADVCCLRGAPKTIOTSPVRSIFVMEPAIG 1126
Db 1062 dehagivpaairpelapiswvthrcpadvccclrgaypmqtsrvlrsifvmepegav 1121
Qy 1127 OKLVYTOAKKANPGAIYVHEAGATFETTTIATADANGLIOSSRAHAYALTRHTEKC 1186
Db 1122 oklvftgaakaanpgsvtcheagatyetltlatadargllqgsrahaivaltrhtekc 1181


```

QY 827 YORFPEAFYTERIMEGLAATLTPRPITIHAVDPYREONKRLAANRETCSRGT 886
DB 822 YQYPSAfaasIvmDgaayLlprplihavpdyrlhmpkrlaearcscrlgta 881
QY 887 AYPPLSGIYQVPSLSFPAWERNHRPDELVTEPAAMWFEANRPAQPLVTEETART 946
DB 882 aypllgtgIyqvpIqpsfawernhrgdelVlylpelaarvfeanrprtlrltledvalt 941
QY 947 ANLAIEIDATEVGRACACCTISPGIVHYQFTAGVPGSKSRISIQGDVYVVVPTRELR 1006
DB 942 anlaieIdatdygracagcgrvtgpyvvyqfIagvpgsksrsltgadvvvyvptreIr 1001
QY 1007 NSMRRRGFAPFTPTAARTIGRRVVIDEAPSLPHLLHLHQRASSVHLGPNQIPAL 1066
DB 1002 nsmrRgfapftptAartIGrrvVIdEapSLphlllHmqraatvhlGpnhlpal 1061
QY 1067 DEHAGLVPAIRPELAPTSMMXVTHRCPADVCELIGAYPKIQTTSRVLSLFWNEPAIG 1126
DB 1062 dtehgllfpairpdlgptscwvthrwpadvcslirgajpmldtcsrvlsIfwgepavg 1121
QY 1127 QKLVYTOAAKAAANPGAITVHEAGATFTETITATADAGLIQSSRAHVALTRTEKC 1186
DB 1122 qkIvftqaakpanpvsvtvheaagatyettIatadargllqssrahaIvaltrhtkx 1181
QY 1187 VILDAAGLREVGISDVYNNFFLAGEVGXHRPSTIPKGNPQNGTLOAFPPSCQISA 1246
DB 1182 vildagllrlevjgsdaIynnfIlaagelghqpsvIprgpanaanvllaafppscqIs 1241
QY 1247 YHOLABELGHRPAVAAYLPCEPELQGLLYMPQELTVSDSVVFEUTDVIHGRMAAPSO 1306
DB 1242 fhglaeelghrpyvaaylppcepegllYlpqelItcdsvvtfeltdIvhermaapsg 1301
QY 1307 KRAVISTLVGRGRRKLYEAASHDVRESLAFPTIGVQATTCCLYELVEAMVEKGD 1366
DB 1302 kxavlstlvgrYgrrklynaashdvrsIarfIpaIgpqvvtceclYelveamvekgqd 1361
QY 1367 GSAVLELDLCNRDVSRTTFQKCKNFTTGETTAHGRKVGOGISAMSKTFCALGPMFRAT 1426
DB 1362 gsavleldlcndvrsrttfqkcnkfttgetIahgkvvggIsawsktfcaltgpmfrat 1421
QY 1427 EKEIILALPNIIFYGDAEESVFAAASGAGSCMFENDFSEPDSTQNNFSLGLECVME 1486
DB 1422 ekaIlaIalppgyvfygdafdtvIsaavaaakasmvfendfseIdstqnnfslgleaIme 1481
QY 1487 ECGMPQWLIRLYHLVRSAMTLOAPKESLKGFMKXSGEPGTLTLMNTWNNATIAHCYEFR 1546
DB 1482 ecgmpqwlIrlyhlrsawllgqpkesslgfkwkhsgepgtlIwntvwmaavltHcydfr 1541
QY 1547 DFRVAAFKGGDDSVYLCSDVROSNAALTAGCGLKUKVDYRPIGLYAGVAVAGLGTLPP 1606
DB 1542 dfrvaafkGGddsvylcsdvrosnaaAltagcglkukvdYrPIglYagvavaglgalpp 1601
QY 1607 VYFAGRLSEKKNMGPPEAEQRLAVCDLRLGTNNVAVQCVVSRVSVSGVLVHNL 1666
DB 1602 vyfagrlsekNmGPpEaeqrlrIavsdflrkltnvaqmcvsvsrVsvsgvlvhnll 1661
QY 1667 GMLQTTADGKAHFTETIKPVLDTJNSIIQRYE 1698
DB 1662 gmlgavadvqkahtesavkpyldtjnsIICrve 1693

```

```

KW HCV: E. coli strain BB4; ATCC deposit number 67717; liver disease.
XX Enterically transmitted non-A, non-B hepatitis virus.
OS WO9115603-A.
XX 17-OCT-1991.
PD 05-APR-1991: 91WO-US02368.
XX 05-APR-1991: 90US-0505888.
PR 05-APR-1990: 90US-0505888.
XX (GENE-) GENELABS INC.
PA (USSH ) US DEPT HEALTH & HUMAN.
XX Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;
PI Fry KE;
XX WPI: 1991-325242/44.
DR N-PSDB: Q14410.
XX New viral proteins from non A-non B hepatitis agent - used to
PT treat and prevent enterically-transmitted non-A non-B hepatitis
PT virus
XX Claim 3; Page 9; 117pp; English.
XX A positive clone ET1.1 was identified in a library prepared from
CC bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.
CC Both strands of ET1.1 were sequenced. One was designated the
CC "forward" strand because of statistical similarities to known
CC proteins and because the forward sequence is known to be
CC predominantly protein-encoding. Of the three possible reading frames
CC only the first (ORF 1) is uninterrupted; the other two include
CC many termination codons. See R14616 and R14617.
XX
SQ Sequence 431 AA:
Query Match 22.8%; Score 2057; DB 12; Length 431;
Best Local Similarity 88.4%; Pred. No. 2.1e-159;
Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;
QY 1257 RRPVAAVLPCEPELQGLLYMPQELTVSDSVVFEUTDVIHGRMAAPSO RKAVALSTLVG 1316
DB 1 rpvvaavlpcepegllYlpqelItcdsvvtfeltdIvhermaapsgrkavlstlvG 60
QY 1317 RYGRRTKLYEAASHDVRESLAFPTIGVQATTCCLYELVEAMVEKGDGSAVLELDLC 1376
DB 61 rygrtklynaashdvrsIarfIpaIgpqvvtceclYelveamvekgdgsavleldlc 120
QY 1377 NRDVSRITTFQKCKNFTTGETTAHGRKVGOGISAMSKTFCALGPMFRATIEKILALPP 1436
DB 12 nrvsrIttfqkcnkfttgetIahgkvvggIsawsktfcaltgpmfratekaIallpp 180
QY 1437 NITFYGAEESVFAAASGAGSCMFENDFSEPDSTQNNFSLGLECVMECEGMPQWLIR 1496
DB 181 nitfygaedfvtIsaavaaakasmvfendfseIdstqnnfslgleaImecgmpqwlIr 240
QY 1497 LVHLVRSAMTLOAPKESLKGFMKXSGEPGTLTLMNTWNNATIAHCYEFDFRVAAFKGD 1556
DB 241 lvhlrsawllgqpkesslgfkwkhsgepgtlIwntvwmaavltHcydfrIvaakgd 300
QY 1557 DSVYLCSDVROSNAALTAGCGLKUKVDYRPIGLYAGVAVAGLGTLPPVYFAGRLSE 1616
DB 301 dsylvcsvrosnaaAltagcglkukvdYrPIglYagvavaglgalppvyfagrlte 360
QY 1617 KMGPGPERAEQRLAVCDLRLGTNNVAVQCVVSRVSVSGVLVHNLGMLQTTADGK 1676
DB 361 kmgpgpereaeqrlrIavsdflrkltnvaqmcvsvsrVsvsgvlvhnllgmIqavadvk 420
QY 1677 AHFTETIKPVL 1687
DB 1677 AHFTETIKPVL 1687

```

enterically transmitted non-A, non-B hepatitis virus; hepatitis C;

XX Sequence 431 AA:

Query Match 22.8%; Score 2057; DB 19; Length 431;
 Best Local Similarity 88.4%; Pred. No. 2.1e-159;
 Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

DB 1257 RPAVAALVPCPELEOGLLYMPQELTVSDVLFVETLIDVHCMAAPSQRKAVSTLVG 1316
 1 rpvvaavlppeleqglllypgeitcdsvtfeltdvnormaapqrxavistlvq 60

DB 1317 RYGRRTKLYEAHSDVRESLAFIPITGPVQATTCETLYELVEMVKEGDSAVLELDC 1376
 61 rygrtklynaashsdvreslaripalipgvyvtcellyelvaemvekqdgsgavleldic 120

DB 1377 NRDVSRITFFQKXCNKFTTGTTAHGKVGQISAMSKTFCALFGPWFRAIEKEITALLPP 1436
 121 nrdvsritffqkcnkfttggtiahgkvqgisamsktfcalfgpwfrakailallppq 180

DB 1437 NIFYGDVAYEESVFAAASGAGSCMVFENDFSEPDSTQNNFSLGLECVMEECGMPQWLIR 1496
 181 nifygdavfddvfaaasgagscmvfendfsepdstqnnfslglecvmeeccmpqwlir 240

DB 1497 LYHLVRSAMTLPAPKESLKGFMKHSGEPTLLMNTVMNAITAHCEFRDRFVAAFKGD 1556
 241 lyhlvrsamtlpapkreslkgfmkhsgeptllmntvmmavltthydrtdrfvaafkgd 300

DB 1557 DSVVLCSDYRQSRNAALTAGCGLKLVDRPIGLYAGVVAPGLGTLDPDVFAPAGRLSE 1616
 301 dsvvlcsdyrqsRNAALTAGCGLKLVDRPIGLYAGVVAPGALPDVVFAPAGRLSE 360

DB 1617 KMMGPPEPRAEQLRLAVCDLRLGTLNVAQVCVDSRVYVSGVLVHNLIGMLQTIADCK 1676
 361 kmvgpperaeqrlrlavsdflrkltnvaqmcvdsrvyvspglvnhnligmqivadgk 420

DB 1677 AHFTETIKPVL 1687
 421 ahftesvkpvl 431

RESULT 15
 B24118
 ID B24118 standard; Protein: 431 AA.
 XX B24118;
 AC B24118;
 DT 29-JAN-2001 (first entry)
 DE Hepatitis E virus protein sequence SEQ ID NO:2.
 XX Hepatitis E virus; HEV; non-A/non-B viral hepatitis; viral protein;
 KW immunological; diagnosis; hepatitis; infection; identification;
 KM detection; immunoreactive; hepatotropic; antinflammatory; virucide;
 KM vaccine; antiviral; antigenic; antibody; antigen.
 XX Hepatitis E virus.
 OS US6120988-A.
 PN US6120988-A.
 XX 19-SEP-2000.
 PD 07-JUN-1995; 95US-0478507.
 PF 25-JUL-1994; 94US-0279823.
 PR 05-APR-1991; 91US-0681078.
 PR 17-JUN-1988; 88US-0208997.
 PR 11-APR-1989; 89US-0336672.
 PR 16-JUN-1989; 89US-0367486.
 PR 13-OCT-1989; 89US-0420921.
 PR 05-APR-1990; 90US-0505888.
 XX (GENE-) GENELABS TECHNOLOGIES INC.
 PA

PA (USCO) US GOVERNMENT.
 XX Yarbough PO, Krawczynski KZ, Fry KE, Bradley DW, Tam A, Reyes GR;
 PI WPI; 2000-593712/56.
 DR N-PSDB; A99255.
 XX

PT Identifying recombinant antigen immunoreactive with antibody induced by
 PT hepatitis E virus (HEV), for detecting HEV infection, comprises
 PT immunoreacting a polypeptide from an HEV genome with an HEV-positive
 PT antiserum -
 XX

PS Claim 6: Column 43-46; 46pp: English.

XX The present invention describes a method for identifying a recombinant
 CC antigen immunoreactive with a hepatitis E virus (HEV)-induced antibody.
 CC The method comprises producing a polypeptide derived from an HEV genome,
 CC immunoreacting the polypeptide with an HEV-positive antiserum and
 CC selecting the polypeptide as a recombinant antigen if the polypeptide
 CC reacts with the HEV-positive antiserum. The method is useful for
 CC identifying recombinant antigen immunoreactive with antibody induced by
 CC HEV. The enterically transmitted non-A/non-B hepatitis (ET-NANB) (also
 CC known as HEV)-specific fragments are useful for identifying ET-NANB-
 CC derived cDNAs, which contain additional sequence information, as primers
 CC for detecting ET-NANB viral genomic material in a patient sample, for
 CC the synthesis of polypeptides for use in immunoassays, and for
 CC identifying similar antigenic regions encoded by related viral strains,
 CC e.g. Burmese strain. The antigens are especially useful in the
 CC preparation of vaccine against ET-NANB infection. These antigens may
 CC further be used to prepare antibodies to ET-NANB virus particles for
 CC use directly as antiviral agents, and to produce antiserum designed for
 CC pre- or post-exposure prophylaxis. The present sequence represents a
 CC specifically claimed HEV protein sequence for use in the present
 CC invention.
 CC

XX Sequence 431 AA:

Query Match 22.8%; Score 2057; DB 21; Length 431;
 Best Local Similarity 88.4%; Pred. No. 2.1e-159;
 Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

DB 1257 RPAVAALVPCPELEOGLLYMPQELTVSDVLFVETLIDVHCMAAPSQRKAVSTLVG 1316
 1 rpvvaavlppeleqglllypgeitcdsvtfeltdvnormaapqrxavistlvq 60

DB 1317 RYGRRTKLYEAHSDVRESLAFIPITGPVQATTCETLYELVEMVKEGDSAVLELDC 1376
 61 rygrtklynaashsdvreslaripalipgvyvtcellyelvaemvekqdgsgavleldic 120

DB 1377 NRDVSRITFFQKXCNKFTTGTTAHGKVGQISAMSKTFCALFGPWFRAIEKEITALLPP 1436
 121 nrdvsritffqkcnkfttggtiahgkvqgisamsktfcalfgpwfrakailallppq 180

DB 1437 NIFYGDVAYEESVFAAASGAGSCMVFENDFSEPDSTQNNFSLGLECVMEECGMPQWLIR 1496
 181 nifygdavfddvfaaasgagscmvfendfsepdstqnnfslglecvmeeccmpqwlir 240

DB 1497 LYHLVRSAMTLPAPKESLKGFMKHSGEPTLLMNTVMNAITAHCEFRDRFVAAFKGD 1556
 241 lyhlvrsamtlpapkreslkgfmkhsgeptllmntvmmavltthydrtdrfvaafkgd 300

DB 1557 DSVVLCSDYRQSRNAALTAGCGLKLVDRPIGLYAGVVAPGLGTLDPDVFAPAGRLSE 1616
 301 dsvvlcsdyrqsRNAALTAGCGLKLVDRPIGLYAGVVAPGALPDVVFAPAGRLSE 360

DB 1617 KMMGPPEPRAEQLRLAVCDLRLGTLNVAQVCVDSRVYVSGVLVHNLIGMLQTIADCK 1676
 361 kmvgpperaeqrlrlavsdflrkltnvaqmcvdsrvyvspglvnhnligmqivadgk 420

DB 1677 AHFTETIKPVL 1687
 421 ahftesvkpvl 431

Search completed: May 30, 2001, 16:09:22
Job time: 199 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:09:23 ; Search time 70.93 Seconds

(without alignments)
2805.851 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGTTTAEQAALAAANSALA.....FTETIKPVLDLTNSIQRVE 1698

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPTREMBL_15:*
2: SP_archaea:*
3: SP_bacteria:*
4: SP_fungi:*
5: SP_human:*
6: SP_invertebrate:*
7: SP_mammal:*
8: SP_mhbc:*
9: SP_organelle:*
10: SP_phage:*
11: SP_plant:*
12: SP_prodent:*
13: SP_unclassified:*
14: SP_vertebrate:*
15: SP_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9001	99.9	1698	14	09YLRL3
2	8812.5	97.8	1708	14	09YK10
3	8780.5	97.4	1708	14	09YLR1
4	7730	85.8	1707	14	09YLV9
5	7468	82.9	1693	14	069410
6	7457	82.7	1693	14	09WML5
7	7456	82.7	1693	14	089444
8	7424	82.4	1693	14	081876
9	7412	82.2	1693	14	081862
10	7333	81.4	1693	14	09WC28
11	7329	81.3	1693	14	039221
12	7304	81.0	1693	14	081344
13	7301	81.0	1693	14	069418
14	3444	38.2	727	14	081873
15	2520	28.0	479	14	071144
16	2365	26.2	452	14	071145
17	2284	25.3	487	14	081869
18	1821	20.2	491	14	09YML2
19	1368	15.2	290	14	081875

20	1183	13.1	245	14	081868	081868 hepatitis e
21	1081	12.0	225	14	09WML5	09WML5 hepatitis e
22	1074	11.9	217	14	081864	081864 hepatitis e
23	1025	11.4	212	14	09YML1	09YML1 hepatitis e
24	988	11.0	210	14	092745	092745 hepatitis e
25	988	11.0	210	14	092746	092746 hepatitis e
26	988	11.0	210	14	092747	092747 hepatitis e
27	988	11.0	210	14	092748	092748 hepatitis e
28	988	11.0	210	14	092749	092749 hepatitis e
29	981	10.9	210	14	056046	056046 hepatitis e
30	913	10.1	193	14	09WML6	09WML6 hepatitis e
31	884	9.8	182	14	081863	081863 hepatitis e
32	769	8.5	158	14	09W952	09W952 hepatitis e
33	763	8.5	152	14	081874	081874 hepatitis e
34	709	7.9	134	14	09YJ14	09YJ14 swine hepat
35	692	7.7	158	14	081867	081867 hepatitis e
36	666	7.4	136	14	09W8P3	09W8P3 hepatitis e
37	665	7.4	136	14	09W8P3	09W8P3 hepatitis e
38	658	7.3	136	14	09W8P3	09W8P3 hepatitis e
39	654	7.3	136	14	09W8P3	09W8P3 hepatitis e
40	646	7.2	136	14	09W8P3	09W8P3 hepatitis e
41	643	7.1	123	14	09QRR7	09QRR7 hepatitis e
42	639	7.1	123	14	09QRR6	09QRR6 hepatitis e
43	638	7.1	123	14	09QRR8	09QRR8 hepatitis e
44	618	6.9	160	14	081865	081865 hepatitis e
45	514	5.7	1704	14	067724	067724 helicoverpa

ALIGNMENTS

RESULT 1
ID 09YLRL3 PRELIMINARY: PRT: 1698 AA.
AC 09YLRL3:
DT 01-MAY-1999 (TREMBLrel. 10, Created).
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE POLYPROTEIN (FRAGMENT).
GN ORF1.
OS Hepatitis E virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US1;
RX MEDLINE=98178637; PubMed=9519822;
RA Schlauder G.G., Dawson G.J., Erker J.C., Knigge M.F., Smalley D.L., Rosenblatt J.E., Desai S.M., Mushawar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis E virus isolated from a patient with acute hepatitis reported in the United States.";
RT J. Gen. Virol. 79:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US1;
RX MEDLINE=99190429; PubMed=10092008;
RA Erker J.C., Desai S.M., Schlauder G.G., Dawson G.J., Mushawar I.K.;
RT "A hepatitis E virus variant from the United States: molecular characterization and transmission in cynomolgus macaques.";
RT J. Gen. Virol. 80:681-690(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US1;
RX Schlauder G.G., Erker J.C., Dawson G.J., Desai S.M., Knigge M.F., Kwo P.Y., Smalley D.L., Rosenblatt J.E., Mushawar I.K.;
RT Submitted (NOV-1997) to the EMBL/Genbank/DBD databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US1;
RX Erker J.C., Schlauder G.G., Dawson G.J., Desai S.M., Mushawar I.K.;
RT Submitted (APR-1998) to the EMBL/Genbank/DBD databases.
DR EMBL: AF060668; AAD15812.1; -.

DR INTERPRO: IPR000606; -
 DR INTERPRO: IPR002588; -
 DR INTERPRO: IPR002589; -
 DR PRAM: PF01443; Viral_helicase1, 1.
 DR PRAM: PF01660; Vmehyltransf, 1.
 DR PRAM: PF01661; DUF27, 1.
 DR Polypotein.
 FT NON_TER
 SQ SEQUENCE 1698 AA; 186069 MM; 10670601150F64EA CRC64;

Query Match 99.98; Score 9001; DB 14; Length 1698;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PGITTAEOAALANANALANAVVVRPLSVOTETILINIMOPQOLVFRPEVLNHHIOR 60
 DB 1 PGITTAEOAALANANALANAVVVRPLSVOTETILINIMOPQOLVFRPEVLNHHIOR 60

QY 61 VINHEBOYCRARAGRCLEVGAPHSINDNPNVLRHCFELRPVGRDVRWYSAPTRGPAN 120
 DB 61 VINHEBOYCRARAGRCLEVGAPHSINDNPNVLRHCFELRPVGRDVRWYSAPTRGPAN 120

QY 121 CRNSALGLPPADRTYCFDFGSCAPAEETGVALLSLHDLMPADVAEMARHGXTRLYAA 180
 DB 121 CRNSALGLPPADRTYCFDFGSCAPAEETGVALLSLHDLMPADVAEMARHGXTRLYAA 180

QY 181 LHPPEVLLPGGYHTSYLLIHGDRAVYVEGDTSGAGNHVSIIRAMIRTKIYGDH 240
 DB 181 LHPPEVLLPGGYHTSYLLIHGDRAVYVEGDTSGAGNHVSIIRAMIRTKIYGDH 240

QY 241 PLYEVRRAIGCHFVLLTLAEPSPMPVYPRSTEVYVRSIFGPGSPSLFSPSASTK 300
 DB 241 PLYEVRRAIGCHFVLLTLAEPSPMPVYPRSTEVYVRSIFGPGSPSLFSPSASTK 300

QY 301 STEFAVNVHIMDRMLTGATLDQAFCCSRMTLRLGISKVTYVGAALVANGMASDAL 360
 DB 301 STEFAVNVHIMDRMLTGATLDQAFCCSRMTLRLGISKVTYVGAALVANGMASDAL 360

QY 361 TAXITAAVLTICHORYLRTQALISKMRRLGVEHAKFTIRLSWLFESGSDYIPGHOLO 420
 DB 361 TAXITAAVLTICHORYLRTQALISKMRRLGVEHAKFTIRLSWLFESGSDYIPGHOLO 420

QY 421 FYACORRWLSAGFLHDPRLVLFEDSVPCRCRFLTKKYAGKFCGCMWLGOCCTCLEPAE 480
 DB 421 FYACORRWLSAGFLHDPRLVLFEDSVPCRCRFLTKKYAGKFCGCMWLGOCCTCLEPAE 480

QY 481 GLVGDHNDHEAYEGSVDAEPRAHLDVSGTYAVHGHOLEALYRALVPODIAARASRLT 540
 DB 481 GLVGDHNDHEAYEGSVDAEPRAHLDVSGTYAVHGHOLEALYRALVPODIAARASRLT 540

QY 541 ATVELVASPRLRECRYLGNKTFRTYVVDGAHLEANGPEQYVLSFDASROSMGASHSLT 600
 DB 541 ATVELVASPRLRECRYLGNKTFRTYVVDGAHLEANGPEQYVLSFDASROSMGASHSLT 600

QY 601 YELTPAGLOVRISNGLDCTATPPPGGAPSAAGEVAAFCSALYRNRFTQRHSLTGLW 660
 DB 601 YELTPAGLOVRISNGLDCTATPPPGGAPSAAGEVAAFCSALYRNRFTQRHSLTGLW 660

QY 661 LHPBGLIGITPPSPGHIWESANPFCGEGTLYRTWSTSGFSSPSPPEAAAPMAATPG 720
 DB 661 LHPBGLIGITPPSPGHIWESANPFCGEGTLYRTWSTSGFSSPSPPEAAAPMAATPG 720

QY 721 LPHSTPVSIDWLPPESEFQVDAAPVAPAPAGLPGVVLTPPPPVHKSIPSPS 780
 DB 721 LPHSTPVSIDWLPPESEFQVDAAPVAPAPAGLPGVVLTPPPPVHKSIPSPS 780

QY 781 RNRRLTYTYDGAKKVYAGSLFESDCMLVNASNGHRPGGGLCHAFYQRPPEAFYPTFT 840
 DB 781 RNRRLTYTYDGAKKVYAGSLFESDCMLVNASNGHRPGGGLCHAFYQRPPEAFYPTFT 840

QY 841 MRGLAAYTLTPPIIHAVAPDVVEQNPRLAAYETOSRGTAYPLLSGITVYVPV 900
 DB 841 MRGLAAYTLTPPIIHAVAPDVVEQNPRLAAYETOSRGTAYPLLSGITVYVPV 900

DB 841 MRGLAAYTLTPPIIHAVAPDVVEQNPRLAAYETOSRGTAYPLLSGITVYVPV 900
 QY 901 SLSPDAMERNHRHPCDELYLTPPEANMEFANKPAPQVLTITTEDARTANLLEIDAAREVG 960
 DB 901 SLSPDAMERNHRHPCDELYLTPPEANMEFANKPAPQVLTITTEDARTANLLEIDAAREVG 960

QY 961 RACAGCTISPGIVHYQFTAGVPGSGKSRSLIOGQDVVDVVPVTRRLSMRRRGFAFTPH 1020
 DB 961 RACAGCTISPGIVHYQFTAGVPGSGKSRSLIOGQDVVDVVPVTRRLSMRRRGFAFTPH 1020

QY 1021 TAARVTIGRRVVIDEAPSLPHLLLMHORAASSVHLIGDPNOIPADIDFHAGLVPAIRPE 1080
 DB 1021 TAARVTIGRRVVIDEAPSLPHLLLMHORAASSVHLIGDPNOIPADIDFHAGLVPAIRPE 1080

QY 1081 LAPTSMXVTHRCPADVCCELIRGAYPKIOTTSTRLSLFNNNEPARGOKLYXTOAKKAMP 1140
 DB 1081 LAPTSMXVTHRCPADVCCELIRGAYPKIOTTSTRLSLFNNNEPARGOKLYXTOAKKAMP 1140

QY 1141 GAITVHEAGATFETIITATDARGLIOSSRAHAIVALTFRHEKCVITLDPGLPREVG1 1200
 DB 1141 GAITVHEAGATFETIITATDARGLIOSSRAHAIVALTFRHEKCVITLDPGLPREVG1 1200

QY 1201 SDYIVNNEFLAGEVGHRSVIRPGNPDONLGLQAFPPSCQISAYHOLAELGHRPAP 1260
 DB 1201 SDYIVNNEFLAGEVGHRSVIRPGNPDONLGLQAFPPSCQISAYHOLAELGHRPAP 1260

QY 1261 VAAVLPCCPEEGGLYMPDELTVSDSVLFELTDIVHCHMAAPSOBKAVLSTLYGRYGR 1320
 DB 1261 VAAVLPCCPEEGGLYMPDELTVSDSVLFELTDIVHCHMAAPSOBKAVLSTLYGRYGR 1320

QY 1321 RTKLYEAHSDVRESLARFTPTIGPVQATTCELYELVEAMVERGODGSAAVLELDLCNRDV 1380
 DB 1321 RTKLYEAHSDVRESLARFTPTIGPVQATTCELYELVEAMVERGODGSAAVLELDLCNRDV 1380

QY 1381 SRITFPOKXCNKFTTGTTIAGKVGOGISAMSKTFCLFQPMWRALKEKTLALLPPIFY 1440
 DB 1381 SRITFPOKXCNKFTTGTTIAGKVGOGISAMSKTFCLFQPMWRALKEKTLALLPPIFY 1440

QY 1441 GDAVEESVFAAASGAGSCAVFENDSEPDSTONNFSLGLECYVMEBCGPMWLIRLYHL 1500
 DB 1441 GDAVEESVFAAASGAGSCAVFENDSEPDSTONNFSLGLECYVMEBCGPMWLIRLYHL 1500

QY 1501 VRSAMWIIQAPKESLKGWKKHSGEPGTLIMNTYWNMAIITHCYEFRDFRVAARKGDSDV 1560
 DB 1501 VRSAMWIIQAPKESLKGWKKHSGEPGTLIMNTYWNMAIITHCYEFRDFRVAARKGDSDV 1560

QY 1561 LCSDYROSRAAALLIAGCGLKLKVDYRPGITLYAGVVVAPGLGTLPDVVRFRAGRLSEKNMG 1620
 DB 1561 LCSDYROSRAAALLIAGCGLKLKVDYRPGITLYAGVVVAPGLGTLPDVVRFRAGRLSEKNMG 1620

QY 1621 PGBERAEOLRLAVCDFLRGLTNNVAQVCVDVVSRYVGPGLVHNLIGMLQTIADGKAHFT 1680
 DB 1621 PGBERAEOLRLAVCDFLRGLTNNVAQVCVDVVSRYVGPGLVHNLIGMLQTIADGKAHFT 1680

QY 1681 ETIKPVLDLTNSITQRYE 1698
 DB 1681 ETIKPVLDLTNSITQRYE 1698

RESULT 2
 Q9YK10 PRELIMINARY; PRT; 1708 AA.
 AC Q9YK10;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE NONSTRUCTURAL POLYPEPTIDE.
 OS WINE HEPACTILIS E VIRUS.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_TaxID=63421;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MENG;

RX MEDLINE=97420774; PubMed=9275216;
 RA Meng X.J., Purcell R.H., Halbur P.G., Lehman J.R., Webb D.M.,
 RA Tsareva T.S., Haynes J.S., Thacker B.J., Emerson S.U.;
 RT "A novel virus in swine is closely related to the human hepatitis E
 RT virus";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9860-9865(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MENG.
 RX MEDLINE=99030877; PubMed=9811705;
 RA Meng X.J., Halbur P.G., Shapiro M.S., Govindarajan S., Bruna J.D.,
 RA Mushahar I.K., Purcell R.H., Emerson S.U.;
 RT "Genetic and experimental evidence for cross-species infection by
 RT swine hepatitis E virus";
 RL J. Virol. 72:9714-9721(1998).
 DR EMBL: AF082843; AAC97208.1; -.
 DR INTERPRO: IPR000606; -.
 DR INTERPRO: IPR002588; -.
 DR INTERPRO: IPR002589; -.
 DR PFAM: PF01443: Viral_helicase1; 1.
 DR PFAM: PF01660: Viral_helicase1; 1.
 DR PFAM: PF01661: DUF27; 1.
 KW Polyprotein.
 FT CHAIN 55 237 METHYLTRANSFERASE.
 FT CHAIN 975 1219 HELICASE.
 FT CHAIN 1222 1708 RNA-DIRECTED RNA POLYMERASE.
 SQ SEQUENCE 1708 AA; 187406 MW; 7444E52DCD616130 CRC64;

Query Match 97.8%; Score 8812.5; DB 14; Length 1708;
 Best Local Similarity 97.8%; Pred. No. 0;
 Matches 1661; Conservative 7; Mismatches 30; Indels 1; Gaps 1;

QY 1 PGTTAIEQALAAANSALANAVVBPFLSRVOTELINLMOPRODVFREVLNHPHIO
 DB 10 PGTTAIEQALAAANSALANAVVBPFLSRVOTELINLMOPRODVFREVLNHPHIO
 QY 61 VHNLEQYCRARAGCLVGAHPRISINPNVNLHRCFLRPVGRDQVQWTSATRGPAAN
 DB 70 AHNLEQYCRARAGCLVGAHPRISINPNVNLHRCFLRPVGRDQVQWTSATRGPAAN
 QY 121 CRRSALRGLPADRYTCFPGFSKCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA
 DB 130 CRRSALRGLPADRYTCFPGFSKCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA
 QY 181 LHLPEVLLPGTYHTSYLLIHDGDAVVTYEGDTSAGYNHDVSLIRANIRTKIIVGDH
 DB 190 LHLPEVLLPGTYHTSYLLIHDGDAVVTYEGDTSAGYNHDVSLIRANIRTKIIVGDH
 QY 241 PLVIEVRAIGCHFVLLTLAAPPSPMPYVPRSTEVYVRSIFGFGSPSLFSPACSTK
 DB 250 PLVIEVRAIGCHFVLLTLAAPPSPMPYVPRSTEVYVRSIFGFGSPSLFSPACSTK
 QY 301 STEFHAVPIHMDLMEGATLDQAFCCSRIMTYLRGISYKTVGALVLANEGNASEDAL
 DB 310 STEFHAVPIHMDLMEGATLDQAFCCSRIMTYLRGISYKTVGALVLANEGNASEDAL
 QY 361 TAVITAAVLTICHORYLFTQAIISKMRRLGVEIAOKFTIRLXSWLEKSGRDYIPGRQLO
 DB 370 TAVITAAVLTICHORYLFTQAIISKMRRLGVEIAOKFTIRLXSWLEKSGRDYIPGRQLO
 QY 421 FTAQCRMLISAGFHLDPRLVLFDESYPGCRFTFLKVAAGKFCFPMWLGQECFLPEAP
 DB 430 FTAQCRMLISAGFHLDPRLVLFDESYPGCRFTFLKVAAGKFCFPMWLGQECFLPEAP
 QY 481 GLVGDGHNEAESESEVPAEPBAHLDVSGTAVHGHOLEALYRALNVPODIAARASRLT
 DB 490 GLVGDGHNEAESESEVPAEPBAHLDVSGTAVHGHOLEALYRALNVPODIAARASRLT
 QY 541 AIVELVAPDRILECRTVLGNKTFRTTIVDGAHLLEANGPEQYVLSFPAASRQSGAGSHSLT
 DB 550 AIVELVAPDRILECRTVLGNKTFRTTIVDGAHLLEANGPEQYVLSFPAASRQSGAGSHSLT

QY 601 YELTPAGLOVRISNGIDCTATPPGAPSAPEGEVAAPFASALYRYNFTQRSHSLTGGLM
 DB 610 YELTPAGLOVRISNGIDCTATPPGAPSAPEGEVAAPFASALYRYNFTQRSHSLTGGLM
 QY 661 LHPEGLLGIFPPSPGHIMESANPCGEGTLYTRTSTSGSSDSFSPPEAAMATPG 720
 DB 670 LHPEGLLGIFPPSPGHIMESANPCGEGTLYTRTSTSGSSDSFSPPEAAMATPG 729
 QY 721 LPHSTPEVSDIWLPPPESEFOYDAAPVPAADPAGLPGCVLT-PPPPVHKPSTIPP 779
 DB 730 LPHSTPEVSDIWLPPPESEFOYDAAPVPAADPAGLPGCVLT-PPPPVHKPSTIPP 789
 QY 780 SRRRLTYTPDGAKYVAGSLFESDDCWLVNANPGHRRPGGGLCHAFYORFPAPFPTER 839
 DB 790 SRRRLTYTPDGAKYVAGSLFESDDCWLVNANPGHRRPGGGLCHAFYORFPAPFPTER 849
 QY 840 IMREGLAAYTLPRPIIHAAPDYRYEONPKRELAAYRECSRRGTAAAYLLSGIYQV 899
 DB 850 IMREGLAAYTLPRPIIHAAPDYRYEONPKRELAAYRECSRRGTAAAYLLSGIYQV 909
 QY 900 VLSLSPDAMERNHRPGDELYTEPAANFEANKPAQVLTITEDTARTANALEIDAATEV 959
 DB 910 VLSLSPDAMERNHRPGDELYTEPAANFEANKPAQVLTITEDTARTANALEIDAATEV 969
 QY 960 GRACAGCTSPGIVHOFTAGVPGSGKRSIOGDVDVVVPPRELNSRRRGFAAFTP 1019
 DB 970 GRACAGCTSPGIVHOFTAGVPGSGKRSIOGDVDVVVPPRELNSRRRGFAAFTP 1029
 QY 1020 HTAARVYIGRRVYIDEAPSLPHLLIHMORASSVHLGDPNOIPADIFHAGLVPAIRP 1079
 DB 1030 HTAARVYIGRRVYIDEAPSLPHLLIHMORASSVHLGDPNOIPADIFHAGLVPAIRP 1089
 QY 1080 ELAPTSMXVYTHRCPADVCELIRGAYPKIOTTSRVLSLFWNEBPAIGOKLYTQAAKAA 1139
 DB 1090 ELAPTSMXVYTHRCPADVCELIRGAYPKIOTTSRVLSLFWNEBPAIGOKLYTQAAKAA 1149
 QY 1140 PGATVHEAGCATFETTTIATADARGLIOSSRAHAVALTRHEKCVIADAGLLREVG 1199
 DB 1150 PGATVHEAGCATFETTTIATADARGLIOSSRAHAVALTRHEKCVIADAGLLREVG 1209
 QY 1200 ISDVIVANFELAGGEVGHRRSVIIPRGPNQNGTLOAFPPSCQISAYHQLAEELGRPA 1259
 DB 1210 ISDVIVANFELAGGEVGHRRSVIIPRGPNQNGTLOAFPPSCQISAYHQLAEELGRPA 1269
 QY 1260 PVAAYLPPCELEQGLLYMPQELTVSDSVLVEFLTDIVHCMAAPSQKAVLSTVGRYG 1319
 DB 1270 PVAAYLPPCELEQGLLYMPQELTVSDSVLVEFLTDIVHCMAAPSQKAVLSTVGRYG 1329
 QY 1320 RRTKLYEAASDVRESLARIPTIGPVQATTCELYELVEMAVEKGDGSAVLELDLCNRD 1379
 DB 1330 RRTKLYEAASDVRESLARIPTIGPVQATTCELYELVEMAVEKGDGSAVLELDLCNRD 1389
 QY 1380 VSRITFEOKXCNKFTTGETTAHGVGOGISAMSKTFPCALGPMFRAIEKILLPPNIF 1439
 DB 1390 VSRITFEOKXCNKFTTGETTAHGVGOGISAMSKTFPCALGPMFRAIEKILLPPNIF 1449
 QY 1440 YGDAYEESVFAAASGAGSCMVEFNDSFSTQNNFSLGLECYVMECGMPQMLIRLYH 1499
 DB 1450 YGDAYEESVFAAASGAGSCMVEFNDSFSTQNNFSLGLECYVMECGMPQMLIRLYH 1509
 QY 1500 LVRSANILQAPKESLGFWMKHSGBEGTLLMNTVMNNAIIAHCEYEPDFVAAFKGDVSV 1559
 DB 1510 LVRSANILQAPKESLGFWMKHSGBEGTLLMNTVMNNAIIAHCEYEPDFVAAFKGDVSV 1569
 QY 1560 VLSGSDYROSNNAAALLAGCLLKVYRPIGLAGVVAAGLGLTLPVVPFAGRLSEKNN 1619
 DB 1570 VLSGSDYROSNNAAALLAGCLLKVYRPIGLAGVVAAGLGLTLPVVPFAGRLSEKNN 1629
 QY 1620 GPGPERAEQRLAVCDLRLGTLNVAQVYDVVSVYGVSPGLVHNLIGMLQTTADGKAHF 1679
 DB 1630 GPGPERAEQRLAVCDLRLGTLNVAQVYDVVSVYGVSPGLVHNLIGMLQTTADGKAHF 1689
 QY 1680 TETIKPVLDTNIIQRYVE 1698

```

Db 1690 TETIKPVLDLNTSIORVE 1708

RESULT 3
O9YLRI PRELIMINARY; PRT: 1708 AA.
AC O9YLRI;
DT 01-MAY-1999 (TREMBLREL. 10, Created)
DT 01-MAY-1999 (TREMBLREL. 10, Last sequence update)
DT 01-JUN-2000 (TREMBLREL. 14, Last annotation update)
DE POLYPROTEIN.
GN ORF1.
OC Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HEV-US2;
RX MEDLINE=98178637; PubMed=9519822;
RA Schläuder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F.,
RA Smalley D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis E virus
RT isolated from a patient with acute hepatitis reported in the United
RT States.";
RL J. Gen. Virol. 79:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-HEV-US2;
RX MEDLINE=99190429; PubMed=10092008;
RA Erker J.C., Desai S.M., Schläuder G.G., Dawson G.J., Mushahwar I.K.;
RT "A hepatitis E virus variant from the United States: molecular
RT characterization and transmission in cynomolgus macaques.";
RL J. Gen. Virol. 80:681-690(1999).
DR EMBL; AF060669; AAD15815.1;
DR INTERPRO; IPR000606;
DR INTERPRO; IPR002588;
DR INTERPRO; IPR002589;
DR PRAM; PF01443; Viral_helicase1;
DR PRAM; PF01660; Vmethyltransf;
DR PRAM; PF01661; DUF27;
DR KMW Polypeptidase.
SQ SEQUENCE 1708 AA: 187260 MW: 80255 BEEDSBEEDAS47 CRC64;

Query Match 97.4%; Score 8780.5; DB 14; Length 1708;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1655; Conservative 8; Mismatches 35; Indels 1; Gaps 1;

```

```

QY 361 TAXITAAVLTICHOYRKYLTQAIKGMRLGVEHAOKFTIRLYSMLFEKSGDYIPRGOLQ 420
DB 370 TAVITAAVLTICHOYRKYLTQAIKGMRLGVEHAOKFTIRLYSMLFEKSGDYIPRGOLQ 429
QY 421 FYAOCRRWLSAGFHLXPRXVLDFEDESVPQCRPTFLKKAAGKCCFMRWLGGCTCTGLEPAE 480
DB 430 FYAOCRRWLSAGFHLXPRXVLDFEDESVPQCRPTFLKKAAGKCCFMRWLGGCTCTGLEPAE 489
QY 481 GLVGDHGDNDNAYGSEVDPAEPALDVSCTGYAAVGHOLEALYALANPODIAARASRLT 540
DB 490 GLVGDHGDNDNAYGSEVDPAEPALDVSCTGYAAVGHOLEALYALANPODIAARASRLT 549
QY 541 ATVELVASPDRLECKRTVYLGKNTFRTVVDGAHLEANGPEQYVLSFDASRQSGAGSHSLT 600
DB 550 ATVELVASPDRLECKRTVYLGKNTFRTVVDGAHLEANGPEEYVLSFDASRQSGAGSHSLT 609
QY 601 YELTPAGLQVRISSNGLDCTATFPFGCAPSAAPGEVAFAFCALRYNKFTRHSLTGLW 660
DB 610 YELTPAGLQVRISSNGLDCTATFPFGCAPSAAPGEVAFAFCALRYNKFTRHSLTGLW 669
QY 661 LHPGLGILFPFSPGHIMESANPFCGEGTLYTRTWSTSGSSPSPPEAAPAMATPG 720
DB 670 LHPGLGILFPFSPGHIMESANPFCGEGTLYTRTWSTSGSSPSPPEAAPAMATPG 729
QY 721 LPHSTPPVSDIWLPPSESEFQVDAAPVAPDPAAGLPVYLT-PPPPPVYHKPSLPP 779
DB 730 LPHSTPPVSDIWLPPSESESHVDAASVSPPEAGLTPYLTLP PPPPVYHKPSLPP 789
QY 780 SRNRRLTYTPDGAQKYVAGSLFESDCDMLVNASNPGHRRPGGLCHAFORPEAFYPEF 839
DB 790 PTRRLTYTPDGAQKYVAGSLFESDCDMLVNASNPGHRRPGGLCHAFORPEAFYPEF 849
QY 840 IMREGLAAYTLTPRPIIHAAVADYRVEQNPRLBAAYRETCSRGTAAYPPLLSGITYOV 899
DB 850 IMREGLAAYTLTPRPIIHAAVADYRVEQNPRLBAAYRETCSRGTAAYPPLLSGITYOV 909
QY 900 VLSFSDMERHNRHRRGDELYLEPAAWFEANKPQPVLTITTEDPRTANLLEIDAAFEV 959
DB 910 VLSFSDMERHNRHRRGDELYLEPAAWFEANKPQPVLTITTEDPRTANLLEIDAAFEV 969
QY 960 GRACAGCTISPGIHYOFTAGVPSGKRSIQGQDVVDVVPVTRFLNRSWRRRFAATP 1019
DB 970 GRACAGCTISPGIHYOFTAGVPSGKRSIQGQDVVDVVPVTRFLNRSWRRRFAATP 1029
QY 1020 HTAAVVTIGRRVVIDEAPSLRPHLLLMORASSVHLIGDPNOIPADFEHAGLVPAIRP 1079
DB 1030 HTAAVVTIGRRVVIDEAPSLRPHLLLMORASSVHLIGDPNOIPADFEHAGLVPAIRP 1089
QY 1080 ELAPTSMMXVTHRCRADVCILRGAYPKIQTSVLRSLFENEPALIGOKLYXQGAARKAN 1139
DB 1090 ELAPTSMMXVTHRCRADVCILRGAYPKIQTSVLRSLFENEPALIGOKLYXQGAARKAN 1149
QY 1140 PGATVTHEAGATFTETIITADARGLIOSSAHAIYALRTHREKCYILDAPGLREVG 1199
DB 1150 PGATVTHEAGATFTETIITADARGLIOSSAHAIYALRTHREKCYILDAPGLREVG 1209
QY 1200 ISDVIYNNFLAGEVGNHRPSVIRGNPDONLCTLAFPSPCOISAYHQAELGHRPA 1259
DB 1210 ISDVIYNNFLAGEVGNHRPSVIRGNPDONLCTLAFPSPCOISAYHQAELGHRPA 1269
QY 1260 PVAAVLPPCPLEBGLLXMPBELTVSDSVLYFELTDIYHGMARPQKAVLSTLYGRYG 1319
DB 1270 PVAAVLPPCPLEBGLLXMPBELTVSDSVLYFELTDIYHGMARPQKAVLSTLYGRYG 1329
QY 1320 RRTKLYEAHSDVRESLARFPTIGPVOATCCELYELVEAWEKGGGSAVLELDLCNRD 1379
DB 1330 RRTKLYEAHSDVRESLARFPTIGPVOATCCELYELVEAWEKGGGSAVLELDLCNRD 1389
QY 1380 VSRITTFQKXCNKFTTGETIYAHGVGOGISAMSKTFCALEFPMPRAIEKEITALLPPIIF 1439
DB 1390 VSRITTFQKXCNKFTTGETIYAHGVGOGISAMSKTFCALEFPMPRAIEKEITALLPPIIF 1449
QY 1440 YGDAYEESVFAAASVGAAGSCWVFENFSEPDSTONNFSLGLECYVMEBCGMPWLIRLYH 1499

```



```

Db 1450 YGDAYEESVFAAASGAGSCMFENDESEFDSQNNFSLGLECVMEECGMQMLRLRH 1509
QY 1500 LYRSANILOAPKESLKGFKKHSERGEPTLLMNTVMMAIIANCYERDRERVAAFKDDSV 1559
Db 1510 LYRSANILOAPKESLKGFKKHSERGEPTLLMNTVMMAIIANCYERDRERVAAFKDDSV 1569
QY 1560 VICSODYROSMAALITAGGGLKTKVYRPIGLYAGVVAAGLGLTDDVYRFAGRLESEKN 1619
Db 1570 VICSODYROSMAALITAGGGLKTKVYRPIGLYAGVVAAGLGLTDDVYRFAGRLESEKN 1629
QY 1620 GGPBERAEOBRLAVCDLRLTNVAOCYDVVSRVYVSGPGLVHNILGMLQTTADGKAHF 1679
Db 1630 GGPBERAEOBRLAVCDLRLTNVAOCYDVVSRVYVSGPGLVHNILGMLQTTADGKAHF 1689
QY 1680 TETIKPVDLINSIIORVE 1698
Db 1690 TENIKPVDLINSIIORVE 1708

RESULT 4
Q9IVZ9 PRELIMINARY; PRT: 1707 AA.
AC Q9IVZ9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis E virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TL;
RA Harrison T.J.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TL;
RA Wang Y., Zhang H., Ling R., Li H., Harrison T.J.;
RT "The complete sequence of hepatitis E virus genotype 4 reveals an
RT alternative strategy for translation of open reading frames 2 and 3."
RL J. Gen. Virol. 81:1675-1686(2000).
DR EMBL: AJ272108: CAB83209.1; -.
SQ SEQUENCE 1707 AA; 187296 MW; 2A80F303AB98C4EC CRC64;

Query Match 85.8%; Score 7730; DB 14; Length 1707;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 1444; Conservative 91; Mismatches 140; Indels 46; Gaps 5;

QY 1 PGTITAEQALAAANSALANAVVAPFLSRVOTELLINLMOPROLVFREPVLAMNPIOR 60
Db 10 PGTITAEQALAAANSALANAVVAPFLSRVOTELLINLMOPROLVFREPVLAMNPIOR 69
QY 61 VTHNELQYCRARAGRCLEVGAPRISINDPNVLRHCFLRPGVRDQVRWTSAPTRGPAN 120
Db 70 VTHNELQYCRARAGRCLEVGAPRISINDPNVLRHCFLRPGVRDQVRWTSAPTRGPAN 129
QY 121 CRRSALRGLPADRTYCFDGFSCAFAEFGVALYSLHDLMPADVAEMARHGXTRLYAA 180
Db 130 CRRSALRGLPADRTYCFDGFSCAFAEFGVALYSLHDLMPADVAEMARHGXTRLYAA 189
QY 181 LHLPEVLLPFGTYRTTSTLLHGDRAVYVTEGDTSAGYNHDSILRAMIRTTKTVGDH 240
Db 190 LHLPEVLLPFGTYRTTSTLLHGDRAVYVTEGDTSAGYNHDSILRAMIRTTKTVGDH 249
QY 241 PLVIERVARIGCHFVLLTAAPESPMPVVPYPRSTREYVRSIFEGGSGSPLEPSCSNK 300
Db 250 PLVIERVARIGCHFVLLTAAPESPMPVVPYPRSTREYVRSIFEGGSGSPLEPSCSNK 309
QY 301 STEHAVVHIWDRMLFGATLDDQAFCCSRLMTYLRGISYKTVGALVANEGNASSEDAL 360

```

```

Db 310 STEHAVVHIWDRMLFGATLDDQAFCCSRLMTYLRGISYKTVGALVANEGNASSEDAL 369
QY 361 TAXITAAVYLTIHQRLRTLOAISKMRILGVEHAQXFTIRLYSMLEKSRDIYIPGQLO 420
Db 370 TAXITAAVYLTIHQRLRTLOAISKMRILGVEHAQXFTIRLYSMLEKSRDIYIPGQLO 429
QY 421 FYAQCRRLMSAGHLDPRVLPDESVPCCRRTFLKKYVAGKFCGCFMRLGQECFCELPAE 480
Db 430 FYAQCRRLMSAGHLDPRVLPDESVPCCRRTFLKKYVAGKFCGCFMRLGQECFCELPAE 489
QY 481 GLVGDGHNDNEAEGSEVDPAEPAHLDVDSGTYAVHGHOLEALRYALNVDPDIAARASRLT 540
Db 490 GRVGEQGYDEAFEGSDIDPAEEATVSIAGSYLVTSQLOPLQALGIPSDLAARASRLT 549
QY 541 ATVELVAPDRLECFVLYGNKTERTTVVDAHLEANGPEQYVLSFPAASRMSGAGSHLT 600
Db 550 ATVELVAPDRLECFVLYGNKTERTTVVDAHLEANGPEQYVLSFPAASRMSGAGSHLT 609
QY 601 YELTPAGLOVYRISNGLDCTATFPFGAPSAAGCEVAAPCSALYRNRFTQRSLTGLM 660
Db 610 YELTPAGLOVYRISNGLDCTATFPFGAPSAAGCEVAAPCSALYRNRFTQRSLTGLM 669
QY 661 LPEGLGLIFPPSPGHHMESANPCGEGTLYRTWSTSGFSSDSPPEAADAAMATPG 720
Db 670 YHPEGLVGLFPPSPGHHMESANPCGEGTLYRTWSTSGFSSDSPPEAADAAMATPG 724
QY 721 LPHSTPVSQIWLPPSE--EFQYDAAP-----VPPAPDAGLP----- 758
Db 725 -----PPAEVNTPEVLDALPSEIMEPAPQAPSAAPSPDSVNSFPTS 770
QY 759 -GPVLTLPPEPPVYKPSIPIPPSRNRLLTYTDGAKVYVAGSLFESCDMLVYANSPNGR 817
Db 771 SGAPLAPPAALPVPVTLGSP---RRLLTYTDGSKVYVAGSLFESCDMLVYANSPNGR 826
QY 818 PGGLCHAFYORPEAFYEFITMRREGLAAYTLTPRIIHAVAPDVRVEONPKRLEAYR 877
Db 827 PGGLCHAFYORPEAFYEFITMRREGLAAYTLTPRIIHAVAPDVRVEONPKRLEAYR 886
QY 878 ETC$SRGTAAYPRLIGSIVOVPSLSFDAMERNHRPDELYLTPANMEANKPAOPVL 937
Db 887 ETC$SRGTAAYPRLIGSIVOVPSLSFDAMERNHRPDELYLTPANMEANKPAOPVL 946
QY 938 TTEPDARTNLALIEDAENVRACAGCTISGCIHYOCTAACPVGSGKRSIQOQDVY 997
Db 947 TTEPDARTNLALIEDAENVRACAGCTISGCIHYOCTAACPVGSGKRSIQOQDVY 1006
QY 998 VVVPTRRLNSWRMRGFAFTPHTAARTIGRRVVIDEAPSLRPHLLLMORASSVHL 1057
Db 1007 IVPTRRLNSWRMRGFAFTPHTAARTIGRRVVIDEAPSLRPHLLLMORASSVHL 1066
QY 1058 GPNQIPADIDFENAGLVPAIRPELAPTSWXYVTHRCPADYCELRGAYPKIOTTSLYRS 1117
Db 1067 GPNQIPADIDFENAGLVPAIRPELAPTSWXYVTHRCPADYCELRGAYPKIOTTSLYRS 1126
QY 1118 LFMNEPAIGOKLYTOAANKANPGAITVHQAQATFETTTIITADARGLIOSSRAHY 1177
Db 1127 LFMNEPAIGOKLYTOAANKANPGAITVHQAQATFETTTIITADARGLIOSSRAHY 1186
QY 1178 ALTRHTEKCIIDAPGLREVGISDVIYVNNFPLAGEVGXHRPSVIRPGNPDNLGTLQA 1237
Db 1187 ALTRHTEKCIIDAPGLREVGISDVIYVNNFPLAGEVGXHRPSVIRPGNPDNLGTLQA 1246
QY 1238 FPPSCQISAYHQAELGHRPAVAVALPCCPELBOGLYMPDELYSDSVYFELTDIV 1297
Db 1247 FPPSCQISAYHQAELGHRPAVAVALPCCPELBOGLYMPDELYSDSVYFELTDIV 1306
QY 1298 HCRMAAPSORKAVLSTLVGYGRTKLYEASHDVDESRLARPTIGPYOATCCELYEV 1357
Db 1307 HCRMAAPSORKAVLSTLVGYGRTKLYEASHDVDESRLARPTIGPYOATCCELYEV 1366
QY 1358 EAMVEKGODGSAYLELDLNCNRDYSRTTFQKCNKFTTGTETIAGKVGQGISAMSKTFCA 1417
Db 1367 EAMVEKGODGSAYLELDLNCNRDYSRTTFQKCNKFTTGTETIAGKVGQGISAMSKTFCA 1426

```

QY 1418 LFGFWFAIEKEILLAPNIFYGDAYEESYFAAAGSAGCWENDESEFSDSTONNES 1477
1427 LFGFWFAIEKEILLAPNIFYGDAYEEDYLAAGAPCKFENDESEFSDSTONNES 1486
QY 1478 LGLCEVMECEGMQWILRLHYLRSAWILQAPKESLKGPMKHSRGPSTLLMTVNMMA 1537
1487 LGLCEIIMECEGMQWILRLHYLRSAWILQAPKESLKGPMKHSRGPSTLLMTVNMMA 1546
QY 1538 IIAHCEYDFRFAAFKGDSDSVLCSYRORNAALAGCGLKLYDPRIGLYAGVYV 1597
1547 VIAHCYEFRLKVAFAFGDSDSVLCSYRORNAALAGCGLKLYDPRIGLYAGVYV 1606
QY 1598 APGIGTLPDYVRFAGRLSEKMGWGPPEBAEOLRLAVCDLRLTNVAQCVVSVRYGV 1637
1607 APGIGTLPDYVRFAGRLSEKMGWGPPEBAEOLRLAVCDLRLTNVAQCVVSVRYGV 1666
QY 1658 SPGLVHNLIGLQRIADGKAHFTETIKPVLDLNTSIIOQVE 1698
1667 SPGLVHNLIGLQRIADGKAHFTETIKPVLDLNTSIIOQVE 1707

RESULT 5
069410 PRELIMINARY; PRT: 1693 AA.
AC 069410;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE METHYL TRANSFERASE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
NCBI_TaxID=12461;
RN (1)
RC SEQUENCE FROM N.A.
RP STRAIN=HEV037;
RA Donati M.C., Fagan E.A., Harrison T.J.:
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: X88292; CAAG6936.1; -
DR INTERPRO: IPR000606; -
DR INTERPRO: IPR002588; -
DR INTERPRO: IPR002589; -
DR PRAM: PR01443; Viral_helicase1; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
DR PFAM: PF01661; DUF27; 1.
KW Transferase.
SQ SEQUENCE 1693 AA; 185190 MW; C347C2436DED9B23 CRC64;

Query Match 82.9%; Score 7468; DB 14; Length 1693;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1402; Conservative 109; Mismatches 167; Indels 26; Gaps 7;

QY 1 PGTITAEQAAALAAANALANAVVVRFLSRVQTEILINLMOPQOLVFRREVLMNHQIR 60
10 PGTITAEQAAALAAANALANAVVVRFLSRVQTEILINLMOPQOLVFRREVLMNHQIR 69
QY 61 VINHELEQYCARAGRCLEYGAPHSINDNPNVLRCELRPVGDVORWYSAPTRCPAAN 120
VINHELEQYCARAGRCLEYGAPHSINDNPNVLRCELRPVGDVORWYSAPTRCPAAN 129
Db VINHELEQYCARAGRCLEYGAPHSINDNPNVLRCELRPVGDVORWYSAPTRCPAAN 129
QY 121 CRSSALRGLEPPADRYCEDFSRCAFAETGVALYSLHDLMPADVADAMRHGXTRLYAA 180
130 CRSSALRGLEPPADRYCEDFSRCAFAETGVALYSLHDLMPADVADAMRHGXTRLYAA 189
Db CRSSALRGLEPPADRYCEDFSRCAFAETGVALYSLHDLMPADVADAMRHGXTRLYAA 189
QY 181 LHPPEVLLPRTGYHTTSYLLIHGDRVAVTYEGDTSAGTNHDYSLIRAWTRTKTYGDH 240
190 LHPPEVLLPRTGYHTTSYLLIHGDRVAVTYEGDTSAGTNHDYSLIRAWTRTKTYGDH 249
Db LHPPEVLLPRTGYHTTSYLLIHGDRVAVTYEGDTSAGTNHDYSLIRAWTRTKTYGDH 249
QY 241 PLVTERRAIGCHFVLLTAAPEPSMPYVYPRSTEVYRSIFGPGTSLPFTSCSTK 300
250 PLVTERRAIGCHFVLLTAAPEPSMPYVYPRSTEVYRSIFGPGTSLPFTSCSTK 309
Db PLVTERRAIGCHFVLLTAAPEPSMPYVYPRSTEVYRSIFGPGTSLPFTSCSTK 309

QY 301 STEFAVPHVIMDRMLFGLATLDQAFCCSRMTYLRGISYKVTVGALVANEGWNASDAL 360
310 STEFAVPHVIMDRMLFGLATLDQAFCCSRMTYLRGISYKVTVGALVANEGWNASDAL 369
Db STEFAVPHVIMDRMLFGLATLDQAFCCSRMTYLRGISYKVTVGALVANEGWNASDAL 369
QY 361 TAXITAAVLTICHOHYRLTQAIKGMRLVJEHOKFTTRLYSLFESGSDYIPGQOLE 420
370 TAXITAAVLTICHOHYRLTQAIKGMRLVJEHOKFTTRLYSLFESGSDYIPGQOLE 429
Db TAXITAAVLTICHOHYRLTQAIKGMRLVJEHOKFTTRLYSLFESGSDYIPGQOLE 429
QY 421 FYAOCRRMLISAGFHLDRVLFEDSVPCRCFTFLKKAQKGCQMRMLGQECSTFLPAE 480
430 FYAOCRRMLISAGFHLDRVLFEDSVPCRCFTFLKKAQKGCQMRMLGQECSTFLPAE 489
Db FYAOCRRMLISAGFHLDRVLFEDSVPCRCFTFLKKAQKGCQMRMLGQECSTFLPAE 489
QY 481 GLVGDHNDENAYGSEVDPAEPALHDVSGTYAVHGOLEALYALVPODIARASLT 540
490 GLVGDHNDENAYGSEVDPAEPALHDVSGTYAVHGOLEALYALVPODIARASLT 549
Db GLVGDHNDENAYGSEVDPAEPALHDVSGTYAVHGOLEALYALVPODIARASLT 549
QY 541 ATVELVASPDLRECRVYLGKNTFRTVVDGHLBANGPEOVYLFSDASROSMAAGSHLT 600
550 ATVELVASPDLRECRVYLGKNTFRTVVDGHLBANGPEOVYLFSDASROSMAAGSHLT 609
Db ATVELVASPDLRECRVYLGKNTFRTVVDGHLBANGPEOVYLFSDASROSMAAGSHLT 609
QY 601 YELTPAGILOVRISNGLDCTATPPPGAPSAAPGEVAFAFCALRYNRFTQHSITGLW 660
610 YELTPAGILOVRISNGLDCTATPPPGAPSAAPGEVAFAFCALRYNRFTQHSITGLW 669
Db YELTPAGILOVRISNGLDCTATPPPGAPSAAPGEVAFAFCALRYNRFTQHSITGLW 669
QY 661 LHPGGLGIFPPSPGHSANPFCGEGTLYTRTW-SGFSDDSP-----PEAAPA 714
670 LHPGGLGIFPPSPGHSANPFCGEGTLYTRTW-SGFSDDSP-----PEAAPA 729
Db LHPGGLGIFPPSPGHSANPFCGEGTLYTRTW-SGFSDDSP-----PEAAPA 729
QY 715 MAATPGLPHTSPVSDLMVLRPPSEEPQVDAAPVPAAPDAGLGPVLRPPPPHPK 774
730 MAATPGLPHTSPVSDLMVLRPPSEEPQVDAAPVPAAPDAGLGPVLRPPPPHPK 773
Db MAATPGLPHTSPVSDLMVLRPPSEEPQVDAAPVPAAPDAGLGPVLRPPPPHPK 773
QY 775 SIPESSNRRLTYTPDGAAYAGSLFESDQDWLVNANPGRPGGLCHAFYORPEAF 834
774 SIPESSNRRLTYTPDGAAYAGSLFESDQDWLVNANPGRPGGLCHAFYORPEAF 829
Db SIPESSNRRLTYTPDGAAYAGSLFESDQDWLVNANPGRPGGLCHAFYORPEAF 829
QY 835 YPTEFIMREGIATLTLPRLPIAHAVADRYEONPKLEAAYRTGCRGTAATPLILGSG 894
830 YPTEFIMREGIATLTLPRLPIAHAVADRYEONPKLEAAYRTGCRGTAATPLILGSG 889
Db YPTEFIMREGIATLTLPRLPIAHAVADRYEONPKLEAAYRTGCRGTAATPLILGSG 889
QY 895 IYOVVSLSFDAWERNRPPGDELTLTPEANWFEANPAPCVLTITDRTATVLALEID 954
890 IYOVVSLSFDAWERNRPPGDELTLTPEANWFEANPAPCVLTITDRTATVLALEID 949
Db IYOVVSLSFDAWERNRPPGDELTLTPEANWFEANPAPCVLTITDRTATVLALEID 949
QY 955 AATEVGRACAGCTISPGIVHYOFTAGVPGSGKSRISIOGDVVDVVVYTRRLNSMRRGF 1014
950 AATEVGRACAGCTISPGIVHYOFTAGVPGSGKSRISIOGDVVDVVVYTRRLNSMRRGF 1009
Db AATEVGRACAGCTISPGIVHYOFTAGVPGSGKSRISIOGDVVDVVVYTRRLNSMRRGF 1009
QY 1015 AATPHTAARVTIGRRVVIDAPSLPRLHLLHMQRASSVHLGDDPNOIPALIDEHAGLV 1074
1010 AATPHTAARVTIGRRVVIDAPSLPRLHLLHMQRASSVHLGDDPNOIPALIDEHAGLV 1069
Db AATPHTAARVTIGRRVVIDAPSLPRLHLLHMQRASSVHLGDDPNOIPALIDEHAGLV 1069
QY 1075 PAIRPELAPTSWVHTRCADVCELIRGAVPKIOTSRVLSFENNEPRLIGOKLVYTOA 1134
1070 PAIRPELAPTSWVHTRCADVCELIRGAVPKIOTSRVLSFENNEPRLIGOKLVYTOA 1129
Db PAIRPELAPTSWVHTRCADVCELIRGAVPKIOTSRVLSFENNEPRLIGOKLVYTOA 1129
QY 1135 AKANPAITVNEHOGATFETITATADARGLIOSSRAHAIVALTHTTEKCVILDAAGL 1194
1130 AKANPAITVNEHOGATFETITATADARGLIOSSRAHAIVALTHTTEKCVILDAAGL 1189
Db AKANPAITVNEHOGATFETITATADARGLIOSSRAHAIVALTHTTEKCVILDAAGL 1189
QY 1195 LREVGISDVIVNFFLAGGEVGHRSVPTRGNDONLGLTLOAFPSCQISAYHQLAEEL 1254
1190 LREVGISDVIVNFFLAGGEVGHRSVPTRGNDONLGLTLOAFPSCQISAYHQLAEEL 1249
Db LREVGISDVIVNFFLAGGEVGHRSVPTRGNDONLGLTLOAFPSCQISAYHQLAEEL 1249
QY 1255 GHRPAPVAAVLPCEPELQGLVLPQSLYSDSVLVLELNDIYHCRMAASORKAVLSTL 1314
1250 GHRPAPVAAVLPCEPELQGLVLPQSLYSDSVLVLELNDIYHCRMAASORKAVLSTL 1309
Db GHRPAPVAAVLPCEPELQGLVLPQSLYSDSVLVLELNDIYHCRMAASORKAVLSTL 1309
QY 1315 VGRYGRRTKLYEAHSDVRSLSARFIPTIGPVQATTCELEIVEAMERKQDSGAVLELD 1374
1310 VGRYGRRTKLYEAHSDVRSLSARFIPTIGPVQATTCELEIVEAMERKQDSGAVLELD 1369
Db VGRYGRRTKLYEAHSDVRSLSARFIPTIGPVQATTCELEIVEAMERKQDSGAVLELD 1369
QY 1375 LCNBDVSRITFEOKXCKNFTTGETTIAHGKVGQGISAMSKTFCALFGFWFAIEKEILL 1434

```

Db 1370 LCNKRDVSRITTFROKCNKTTGTTGTHAGKVGCGISAMSTFCALFEPWRAITAKALLALL 1429
Qy 1435 PNIFYGAYEESVFAAAYSGAGSCWYFENDESEFDSTONNFSLECYVMEBCGNPOML 1494
Db 1430 PQVFYGDADFDTVFAAATAAKASWVFENDESEFDSTONNFSLECYVMEBCGNPOML 1489
Qy 1495 IRLYHLIRASAMLIQAKRESLKGFWKHSSEPGTLMTNTYNNALIAHCEYFRFRRAAK 1554
Db 1490 IRLYHLIRASAMLIQAKRESLKGFWKHSSEPGTLMTNTYNNALIAHCEYFRFRRAAK 1549
Qy 1555 GDSVYLCSDYRSMNAALIAAGCGLTKVDYRPIGLYGVVYAPGLGTLPDVYFRAGRL 1614
Db 1550 GDSVYLCSEYRSGPAATVLIAGCGLTKVDYRPIGLYGVVYAPGLGTLPDVYFRAGRL 1609
Qy 1615 SEKNMGPERAEQRLAYCDLRLGTLNTYAQVCVDVYSRVYSGPLVHNLIGMLQTIAD 1674
Db 1610 TEKMGPERAEQRLAYCDLRLGTLNTYAQVCVDVYSRVYSGPLVHNLIGMLQTIAD 1669
Qy 1675 GRAHFTETIKPVLDLTNSTIQRYE 1698
Db 1670 GRAHFTESYKPVLDLTNSTILCRVE 1693

RESULT 6
Q9WLL5 PRELIMINARY: PRT: 1693 AA.
ID Q9WLL5;
AC Q9WLL5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK15/92;
RA MEDLINE=99049628; PubMed=9833882;
RX Gouvea V., Snellings N., Popek M.D., Longer C.F., Innis B.L.;
RT "Hepatitis E virus: complete genome sequence and phylogenetic analysis
of a Nepal isolate."
RT Virus Res. 57:21-26(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TK15/92;
RA Gouvea V.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF051830; AAC97186.1; -
DR INTERPRO: IPR000606; -
DR INTERPRO: IPR002588; -
DR INTERPRO: IPR002589; -
DR PFAM: PF01443; Viral_helicase1.1.
DR PFAM: PF01660; Vmchyltransf.1.
DR PFAM: PF01661; DUF27.1.
KW Polyprotein.
SQ SEQUENCE 1693 AA: 185349 MW: A895ACDIDAEZFBBD CRC64:

Query Match 82.7%; Score 7457; DB 14; Length 1693;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1397; Conservative 110; Mismatches 166; Indels 36; Gaps 6;

```

```

Db 130 CRSSALRGILPADRTYCFDGFSGCSFPPEIGALISLHDMSPDVAEMARRHGCTRLLYA 189
Qy 181 LHLPEVLLPPGYHTTYSYLLHDDGRAVVTYEGDTSAGYNHDVSLIRANIRTKIYGDH 240
Db 190 LHLPEVLLPPGYHTTYSYLLHDDGRAVVTYEGDTSAGYNHDVSLIRANIRTKIYGDH 249
Qy 241 PLVIERVRAIGHFVLLLTAAPESPMPYPRYPRSTREYVYRSIFGCGSGSLSPSACSTK 300
Db 250 PLVIERVRAIGHFVLLLTAAPESPMPYPRYPRSTREYVYRSIFGCGSGSLSPSACSTK 309
Qy 301 STFAHVAHIMDMLMFGATLDDQAFCCSRLMTYLGISYKVYVAGLVANEGNASEDAL 360
Db 310 STFAHVAHIMDMLMFGATLDDQAFCCSRLMTYLGISYKVYVAGLVANEGNASEDAL 369
Qy 361 TAXITAAVLTTCQRYLRQAISKMRRLGVEHAQKFTRLYSWLPFKSGRDYIPGRQLO 420
Db 370 TAXITAAVLTTCQRYLRQAISKMRRLGVEHAQKFTRLYSWLPFKSGRDYIPGRQLO 429
Qy 421 FYAQCRRWLISAGFHLDPRVLYPDESVCRCRTFLKKYAGKFCGCMRLGECCTFLPAE 480
Db 430 FYAQCRRWLISAGFHLDPRVLYPDESAPCHCRVIRKVLSECFPMWMLGECCTFLPAE 489
Qy 481 GLVGDHGDNEAEVGESEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPQDIARASRLT 540
Db 490 GYVGDDGHNEAEVGESEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPQDIARASRLT 549
Qy 541 ATVELVASPDRLCEKRVLYGNKTRFTTYVDGAHLIDANGPEOYVLSFPAASROSGAGSHLT 600
Db 550 ATVELVASPDRLCEKRVLYGNKTRFTTYVDGAHLIDANGPEOYVLSFPAASROSGAGSHLT 609
Qy 601 YELTPAGLOVIRISNGIDCATFPFGAPSAAGAEVAFCSALYRYRFRQHSISLGLW 660
Db 610 YASASAGLEVRYAAGLDHAIIFAPGVSPSTGEVYAFCSALYRYRFRQHSISLGLW 669
Qy 661 LHPEGLIGLPPSPGIMESANPFCEGGLYRTWS-TSGFSSDESP-PEAAPA 714
Db 670 FHPEDGLIGLPPSPGIMESANPFCEGGLYRTWS-TSGFSSDESP-PEAAPA 729
Qy 715 MAATPGLPHSTPPVSDIWLVPPESEFQVDAAPV-PPAPDAGLPGHVLTLPPEPPVHK 773
Db 730 RAATPPL-----AALPPAPDPS--PPFPAPAPPEPAPGT 764
Qy 774 PSIP-----PPSRNRLITYPDGAKYVAGSLFSDCOMLVNANSPGRPGGCHAHAYR 829
Db 765 AGAPALITHOMARRRLFTYPDGSKVYAGSLFSDCOMLVNANSPGRPGGCHAHAYR 824
Qy 830 FPEAFYPTFERIMEGLAAYTLTPRPIIHAAVAPDYVQNPFRLEAAVRETCRSRGTAAAP 889
Db 825 YPASFDASVYMDGAAAYTLTPRPIIHAAVAPDYVQNPFRLEAAVRETCRSRGTAAAP 884
Qy 890 LCGSGLYQVPSLSPDAMERNHRRGDELYLTPRAAMFEANKPAPVLTTEDTARTANL 949
Db 885 LCGSGLYQVPSLSPDAMERNHRRGDELYLTPRAAMFEANKPAPVLTTEDTARTANL 944
Qy 950 ALFIDATFVGRACAGCTIPGIVHYOFTAGVGSKGSRSIOGDVYVVVPPRELNSW 1009
Db 945 ALFIDATFVGRACAGCTIPGIVHYOFTAGVGSKGSRSIOGDVYVVVPPRELNSW 1004
Qy 1010 RRRGFAFTPTHTAARVITGRVVIDEAPSLPHLLILHMRASSVHLIDGPNQIPADFE 1069
Db 1005 RRRGFAFTPTHTAARVITGRVVIDEAPSLPHLLILHMRASSVHLIDGPNQIPADFE 1064
Qy 1070 HAGLVPAIRPELAPTSKMWYTHRCPADVCBLINGANPKIOTSRVLSLFWNEPAIGOKL 1129
Db 1065 HAGLVPAIRPELAPTSKMWYTHRCPADVCBLINGANPKIOTSRVLSLFWNEPAIGOKL 1124
Qy 1130 VTOGAKAANPAGATVTEAGATFTEETITATDARGLIOSSRAHVALTRHREKCVIL 1189
Db 1125 VTOGAKAANPAGATVTEAGATFTEETITATDARGLIOSSRAHVALTRHREKCVIL 1184
Qy 1190 DAPGLREVGISDVYVNNPFLAGGEVGHRRPSVYPRGNPQNLGTLQAFPPSQISAHYQ 1249
Db 1185 DAPGLREVGISDVYVNNPFLAGGEVGHRRPSVYPRGNPQNLGTLQAFPPSQISAHYQ 1244

```

QY 1250 LAELGHPAPVAVLPPCPPELEGLYMPDELTIVSDSVLELTDIYHCRMAAPDSQKA 1309
D 1245 LAELGHPAPVAVLPPCPPELEGLYMPDELTIVSDSVLELTDIYHCRMAAPDSQKA 1304
QY 1310 VLSTLVGVRGRTKLYEAASHDVRESLARFTPTGPGVATCELYELVEANVEKQDSSA 1369
D 1305 VLSTLVGVRGRTKLYEAASHDVRESLARFTPTGPGVATCELYELVEANVEKQDSSA 1364
QY 1370 VLELDICNRDVSRTFFPKXCNKFTTGFTIAGKVGOGISAMSKTFCALFGPMPRAIKE 1429
D 1365 VLELDICNRDVSRTFFPKXCNKFTTGFTIAGKVGOGISAMSKTFCALFGPMPRAIKE 1424
QY 1430 ILALIPNIEFYGDAAVEESVFAAASGASGVFENFSEFSTONNFSILGECVWMECG 1489
D 1425 ILALIPNIEFYGDAAVEESVFAAASGASGVFENFSEFSTONNFSILGECVWMECG 1484
QY 1490 MPOMLIRLYHLIRSAWILQAKESIKGFWKHSGEPGTLNNTYNNALITAHCEFRFR 1549
D 1485 MPOMLIRLYHLIRSAWILQAKESIKGFWKHSGEPGTLNNTYNNALITAHCEFRFR 1544
QY 1550 VAAKGGDSVILCSEYRSPGAVALIAGCGIKLVDPFRTGLYAGVYVAPGLPDVVR 1609
D 1545 VAAKGGDSVILCSEYRSPGAVALIAGCGIKLVDPFRTGLYAGVYVAPGLPDVVR 1604
QY 1610 FAGRLSEKNMGCPGPRAEQRLAYCDLRLTNAQVCVDVSVRYGVSPGLVHNLIGML 1669
D 1605 FAGRLSEKNMGCPGPRAEQRLAYCDLRLTNAQVCVDVSVRYGVSPGLVHNLIGML 1664
QY 1670 OTIADGKAHFTETIKPVLDTLNTSIORYE 1698
D 1665 OTIADGKAHFTETIKPVLDTLNTSIORYE 1693

RESULT 7
089444 PRELIMINARY: PRT: 1693 AA.
AC 089444:
DT 01-NOV-1996 (Tremblrel, 01, Created)
DT 01-NOV-1996 (Tremblrel, 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel, 14, Last annotation update)
DE UNANMED PROTEIN PRODUCT.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI Taxid=12461:
RN [1]
RA YIN S. R., Purcell R. H., Emerson S. U.:
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K52-87;
RX MEDLINE=95176571; PubMed=7871758;
RA YIN S., Purcell R. H., Emerson S. U.:
RT "A new Chinese isolate of hepatitis E virus: comparison with strains
RT recovered from different geographical regions."
RT Virus Genes 9:23-32(1994).
DR EMBL: L25547; AAA91078.1; -
DR EMBL: L25595; AAA65488.1; -
DR INTERPRO: IPR000606; -
DR INTERPRO: IPR002588; -
DR INTERPRO: IPR002589; -
DR PFAM: PF01443; Viral_helicase1; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
DR PFAM: PF01661; DUF27; 1.
SQ SEQUENCE 1693 AA; 185122 MW; 53914B830238D5E CRC64;

Query Match 82.7%; Score 7456; DB 14; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1401; Conservative 108; Mismatches 161; Indels 42; Gaps 8;

QY 1 PGITTAIEQAALAAANSALANAVVVRPFLSVQTEILINLMQPRQVLPREVLMNHPIOR 60

D 10 PGITTAIEQAALAAANSALANAVVVRPFLSVQTEILINLMQPRQVLPREVLMNHPIOR 69
QY 61 VIHNELEQYCARAGRCLEVGAAHPRSINDPNVLRHCTLRPVGRVQWYASAPTRGPAN 120
D 70 VIHNELEQYCARAGRCLEVGAAHPRSINDPNVLRHCTLRPVGRVQWYASAPTRGPAN 129
QY 121 CRRSALRGLPPADRTYCDGFSRCAFAAETGVALYSLHDMPADVAEAMARHGTRLYAA 180
D 130 CRRSALRGLPPADRTYCDGFSRCAFAAETGVALYSLHDMPADVAEAMARHGTRLYAA 189
QY 181 LHLPEVLLPRTYHTTSYLLIHGDRAVVTYEGDTSAGYVHDSILRAMIRTKIYGDH 240
D 190 LHLPEVLLPRTYHTTSYLLIHGDRAVVTYEGDTSAGYVHDSILRAMIRTKIYGDH 249
QY 241 PLVIERVATICHVYLLTAAPEPSMPYVYPPSTEVYVNSIFGPGSPSLFSPSACTK 300
D 250 PLVIERVATICHVYLLTAAPEPSMPYVYPPSTEVYVNSIFGPGSPSLFSPSACTK 309
QY 301 STFAVPAVHMDRLMEGATLDQAFCSSRLMTYLRGISTYKVTYVGAALVANGMNASDAL 360
D 310 STFAVPAVHMDRLMEGATLDQAFCSSRLMTYLRGISTYKVTYVGAALVANGMNASDAL 369
QY 361 TAXITTAAYLTICHOYLTQTAISKGMRLGYEHAOKFTRLYSMLFEKSGDYIPGROLQ 420
D 370 TAXITTAAYLTICHOYLTQTAISKGMRLGYEHAOKFTRLYSMLFEKSGDYIPGROLQ 429
QY 421 FYACRRRLSGFHLDPVLVFEDESVPORCTPLKTKVAGKCCMRMLGOCCTFLPAE 480
D 430 FYACRRRLSGFHLDPVLVFEDESVPORCTPLKTKVAGKCCMRMLGOCCTFLPAE 489
QY 481 GLVGDHGHNEAYGSEVPAEPALHDVSGTYAVHGOLEALYBALVNPODIARASRLT 540
D 490 GLVGDHGHNEAYGSEVPAEPALHDVSGTYAVHGOLEALYBALVNPODIARASRLT 549
QY 541 ATVELVASPDRLCERYLGKNTFTTYVDCAHLNANGPEOYVLSFDAKSGMGAGSHSLT 600
D 550 ATVELVASPDRLCERYLGKNTFTTYVDCAHLNANGPEOYVLSFDAKSGMGAGSHSLT 609
QY 601 YELPAGLOVYISSNGDCATPPPGGAPSAPEVAFAFCALRYRNFQRLHSLTGLW 660
D 610 YELPAGLOVYISSNGDCATPPPGGAPSAPEVAFAFCALRYRNFQRLHSLTGLW 669
QY 661 LHPGLGTFPPSPGHIMESANPFCGGLYTRTWS-----TSGFSSDSPPE 709
D 670 LHPGLGTFPPSPGHIMESANPFCGGLYTRTWS-----TSGFSSDSPPE 729
QY 710 AAAPAMATPGLPHSTPVSIDIVLPPPSSEFOVDAAPVPPAP--DPAGLPGVYVLTTP 766
D 730 AAAPAMATPGLPHSTPVSIDIVLPPPSSEFOVDAAPVPPAP--DPAGLPGVYVLTTP 766
QY 767 PPPVHKRPSIPPSRNRRLTYTPDGAKVYAGSLFESDCDMLVNASNGHPPGGGLCHAF 826
D 767 APATTHQ-----TARHRRLLFTYPDGSKVAFSGLFESTCTMLVNASVNDHRPGELCHAF 821
QY 827 YORPEAFYPTFTIMREGLAAYTLTPRPIIHAAVADYRVEONPRKLEAAYRETCSRGTA 886
D 822 YORPEAFYPTFTIMREGLAAYTLTPRPIIHAAVADYRVEONPRKLEAAYRETCSRGTA 881
QY 887 AYPLLSGTYOVVYSFDMERNHRGDELXLEPAAANFEAKKPAOPVLTITTEDART 946
D 882 AYPLLSGTYOVVYSFDMERNHRGDELXLEPAAANFEAKKPAOPVLTITTEDART 941
QY 947 ANLLEIDAATEVGRACAGCTISPGIYHOFTAGVPSGSKRSIOGDVYVVPVPTREL 1006
D 942 ANLLEIDAATEVGRACAGCTISPGIYHOFTAGVPSGSKRSIOGDVYVVPVPTREL 1001
QY 1007 NSWRRGFAAFTPHTAARVTTGRRVVIDEAPSLPHLLLLHMRASSVHLGDPNOIPAI 1066
D 1002 NSWRRGFAAFTPHTAARVTTGRRVVIDEAPSLPHLLLLHMRASSVHLGDPNOIPAI 1061
QY 1067 DFEHAGLVPAIRPBLATSWXVYTHRCPADVCELIKATYKIQTSYVLSLFPANEPAG 1126

```
DB 1062 DEHAGLVPAIRBDLAPTSMMHVTNRCPADVCELLINGAPMIQOTTSRVLSLFWGEPAVG 1121
QY 1127 OKLVYQAANKANPGALITYHEAGATFTTITATADAGLISSAHAIVALTRTEKC 1186
DB 1122 OKLVYQAANKANPGSVTAEAGATYTTTITATADAGLISSAHAIVALTRTEKC 1181
QY 1187 VIIDAPGLREVGISPVYNNFFLAGGEVGHHPSPVIRPGNPONQITGOAFPSQOISA 1246
DB 1182 VIIDAPGLREVGISDAIYNNFFLAGGELIGHORPSVIRPGNPANDDTLAAPPSCQISA 1241
QY 1247 YHQAELGHRPAPVAAPVPCPELEQGLLXMPQELTVSDVLFELTDIVHCRMAPSQ 1306
DB 1242 FHQAELGHRPAPVAAPVPCPELEQGLLXMPQELTVSDVLFELTDIVHCRMAPSQ 1301
QY 1307 KRAVSTLVGRGRKTKLYEAASHVRESLAPITPTIGVQATTGCLYELVEAMVERGD 1366
DB 1302 KRAVSTLVGRGRKTKLYEAASHVRESLAPITPTIGVQATTGCLYELVEAMVERGD 1361
QY 1367 GSVALELDCNDRVSRITFEOKKCNKFTTGETTAHGKVGSGISAMSKTCALGPMFRAT 1426
DB 1362 GSVALELDCNDRVSRITFEOKKCNKFTTGETTAHGKVGSGISAMSKTCALGPMFRAT 1421
QY 1427 EKELIALLPNTIFYGDAYEESVFAAASGAGSCMVEENFSEPDSTONNESLGLCEVME 1486
DB 1422 EKALIALLPQGYFYGDAPDDYFSAVAAMAKSMVEENFSEPDSTONNESLGLCEAIME 1481
QY 1487 ECGMOWMLRLYLHVASANTLQAPKESLKGFMKHSGEPTGLMNTVWMAITAHCEYR 1546
DB 1482 ECGMOWMLRLYLHVASANTLQAPKESLKGFMKHSGEPTGLMNTVWMAITAHCEYR 1541
QY 1547 DRRVAFAKGDSDVILCSDRQSRNAALJAGCLKLVYRPIGLAGVVAAPGLGTLPP 1606
DB 1542 DLQVAFAKGDSDVILCSDRQSRNAALJAGCLKLVYRPIGLAGVVAAPGLGTLPP 1601
QY 1607 VVRFAGRLEKMMWGPPEBAEQLRLAFCGLRTNVAQCVDVSVRYGVSPGLVHNL 1666
DB 1602 VVRFAGRLEKMMWGPPEBAEQLRLAFCGLRTNVAQCVDVSVRYGVSPGLVHNL 1661
QY 1667 GMLQITADGKAHTTETIKIPVLDLNTSIQRYE 1698
DB 1662 GMLQAVADGKAHTTESVKPVLDTLNTSILSRVE 1693

RESULT 8
OB1876
AC 081876: PRELIMINARY: PRT: 1693 AA.
DB 01-NOV-1996 (TREMBLrel. 01, Created)
DB 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DB 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE COMPLETE GENOME SEQUENCE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UIGH179;
RA Uchida T.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; D11093; BAA01865.1;
DR INTERPRO; IPR000606;
DR INTERPRO; IPR002589;
DR INTERPRO; IPR002589;
DR PFAM; PF01443; Viral_helicase1.1;
DR PFAM; PF01600; Vmethyltransf.1;
DR PFAM; PF01661; DUF27.1;
SQ SEQUENCE 1693 AA; 184839 MW; 4E4222ED55AE0B1 CRC64;
```

Query Match 82.4%; Score 7424; DB 14; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 108; Mismatches 165; Indels 42; Gaps 8;

```
QY 1 PGITTAIEQALAAANSALANAVVPRFLSRVOTELINLMOROLVPRREVLMNHPOR 60
DB 10 PGITTAIEQALAAANSALANAVVPRFLSHQOIELINLMOROLVPRREVLMNHPOR 69
QY 61 VINELEOYCARAGRCLEVGAPHSINDPNVNLHRCFLRVPYRDVQRMWTSAPTRGPAN 120
DB 70 VINELEOYCARAGRCLEVGAPHSINDPNVNLHRCFLRVPYRDVQRMWTSAPTRGPAN 129
QY 121 CRSSALRGLEPRADRYCFDGFSCAPAEATGVALYSLHDIIMPADVAEMARHGXTLYAA 180
DB 130 CRSSALRGLEPRADRYCFDGFSCAPAEATGVALYSLHDIIMPADVAEMARHGXTLYAA 189
QY 181 LHLPEVLLPBGTYHTSYLLIHGDRAVTYGSDISAGNNHVSILRAIFRTKIVGDH 240
DB 190 LHLPEVLLPBGTYHTSYLLIHGDRAVTYGSDISAGNNHVSILRAIFRTKIVGDH 249
QY 241 PLVIEVRALGCHFFVLLTNAPESPMPVYPRSTEVYRSIFGCGSPSLPASCSTK 300
DB 250 PLVIEVRALGCHFFVLLTNAPESPMPVYPRSTEVYRSIFGCGSPSLPASCSTK 309
QY 301 STEHAVPVHIMDLMLPGATLDDQAFCCSLMTYLRGISTKYTVGALVANEGNASEDAL 360
DB 310 STEHAVPVHIMDLMLPGATLDDQAFCCSLMTYLRGISTKYTVGALVANEGNASEDAL 369
QY 361 TAXITAAVLTICHORYLRTQALISGMRLGVEHAOKFITFLYSMLFEKSGROYIPGROL 420
DB 370 TAXITAAVLTICHORYLRTQALISGMRLGVEHAOKFITFLYSMLFEKSGROYIPGROL 429
QY 421 FYAQCRMLTSAGHLDPRVLYVFEDESVCRCRTLTKAKAGFCCEFMRLGECCTFLEPAE 480
DB 430 FYAQCRMLTSAGHLDPRVLYVFEDESVCRCRTLTKAKAGFCCEFMRLGECCTFLEPAE 489
QY 481 GLVGDGHNDNEAVEGSEVDPAPBAHLDVSGTYAVHGHOLEALYRALNPODIAARSLT 540
DB 490 GLVGDGHNDNEAVEGSEVDPAPBAHLDVSGTYAVHGHOLEALYRALNPODIAARSLT 549
QY 541 ATVELVASPRLCRFLYLGKTRTYVVDGAHLNANGPEQVYLSFASRQSMGSHSLT 600
DB 550 ATVELVASPRLCRFLYLGKTRTYVVDGAHLNANGPEQVYLSFASRQSMGSHSLT 609
QY 601 YELTPAGLOVRISNGLDCTATPEPPGAPSAPEVAFAFCALYRYNRFTQRLSLTGLM 660
DB 610 YELTPAGLOVRISNGLDCTATPEPPGAPSAPEVAFAFCALYRYNRFTQRLSLTGLM 669
QY 661 LHPGGLIGTPPSPGHIIMSANPCGEGTLTYRTMS-----TGFSSDSPE 709
DB 670 LHPGGLIGTPPSPGHIIMSANPCGEGTLTYRTMS-----TGFSSDSPE 729
QY 710 AARPMAATPGLHSPRPVSDIWLPPPSSEFQVDAAPVPAP---DPAGLPQGVYLTLP 766
DB 730 AARPMAATPGLHSPRPVSDIWLPPPSSEFQVDAAPVPAP---DPAGLPQGVYLTLP 766
QY 767 PRPVHKKPSIPPSRNRLTYTPDGAKYVAGSLFESDCMLVNASNPGHRRPGGLCHAF 826
DB 767 PRPVHKKPSIPPSRNRLTYTPDGAKYVAGSLFESDCMLVNASNPGHRRPGGLCHAF 821
QY 827 YQFPEAFYPTERIMBGLAAYTLTPRPIIHAVAPDYRVQNPRLKLEAAYRETSRRGTA 886
DB 822 YQFPEAFYPTERIMBGLAAYTLTPRPIIHAVAPDYRVQNPRLKLEAAYRETSRRGTA 881
QY 887 AYPFLGSGIYQVPSLSFQMERNNHPRGDELXLTPEAANFEANKRQPVLTTEDEART 946
DB 882 AYPFLGSGIYQVPSLSFQMERNNHPRGDELXLTPEAANFEANKRQPVLTTEDEART 941
QY 947 ANALEIDATEVGRACAGCTISPGIYHYFTAGVSGSGRSIQOQDVVVVVVPTRELR 1006
DB 942 ANALEIDATEVGRACAGCTISPGIYHYFTAGVSGSGRSIQOQDVVVVVVPTRELR 1001
QY 1007 NSWRRRGFAAFTPTAARTYIGRRVVIDEAPSLPHILLHMQASSVHLLGPNQIPAT 1066
DB 1002 NSWRRRGFAAFTPTAARTYIGRRVVIDEAPSLPHILLHMQASSVHLLGPNQIPAT 1061
QY 1067 DEHAGLVPAIRBDLAPTSMMHVTNRCPADVCELLINGATPKIOTTSRVLSLFWGEPAVG 1126
```

```

|||||
Db 1062 DFEHAGLVPAIRPDLAPTSMMHWHTRCPADVCCELLRGAYPMIQTSTRVLSRFWCEPAVG 1121
QY 1127 OKIYXTOAKANKGATVYHBAOGATFETETITATADARGLIQSSRAHAYALPHTEK 1186
Db 1122 OKLVTOAKANKGATVYHBAOGATFETETITATADARGLIQSSRAHAYALPHTEK 1181
QY 1187 VILDPAGLREVGISDVYVNEFFLAGEGVGHRSVIRGNDDOILGLOAFPPSCQISA 1246
Db 1182 VILDPAGLREVGISDVYVNEFFLAGEGVGHRSVIRGNDDOILGLOAFPPSCQISA 1241
QY 1247 YHOLAEELGHRPAPYAAVLPCEPELGLLMPDELIVSDVLFELTDIVHCRMAPSQ 1306
Db 1242 YHOLAEELGHRPAPYAAVLPCEPELGLLMPDELIVSDVLFELTDIVHCRMAPSQ 1301
QY 1307 RKAVALSTVGRGRTKLYEAAHSDVRESLRFPTIGPVATTCCELVEIYEAWEKGD 1366
Db 1302 RKAVALSTVGRGRTKLYEAAHSDVRESLRFPTIGPVATTCCELVEIYEAWEKGD 1361
QY 1367 GSAAVLELDLCNRDVSRTFFQKXCNKFTTGETIAHGKVGOGISAMSKTFCALFGWFRAI 1426
Db 1362 GSAAVLELDLCNRDVSRTFFQKXCNKFTTGETIAHGKVGOGISAMSKTFCALFGWFRAI 1421
QY 1427 EKELLALPRIFEGDAEESVFAAASGASCAWFEDESEPDSTONNESLGECEYME 1486
Db 1422 EKELLALPRIFEGDAEESVFAAASGASCAWFEDESEPDSTONNESLGECEYME 1481
QY 1487 ECGMPDWLIRLYHVRASAMITQAPKESLKGFKHSGEPGLTNTVNNMAIIAHCEFR 1546
Db 1482 ECGMPDWLIRLYHVRASAMITQAPKESLKGFKHSGEPGLTNTVNNMAIIAHCEFR 1541
QY 1547 DFRVAAEKGDSDSVYLCSDYROSNNAAIIAGCGKLKLVDPYRIGLYAGVVAAPGLGLPD 1606
Db 1542 DLQVAAEFKGDSDSVYLCSDYROSNNAAIIAGCGKLKLVDPYRIGLYAGVVAAPGLGLPD 1601
QY 1607 VVRVAGRLSEKMGKGPRAQOLRLAVCDLRLTNVAGVVDVYRXYGSPGLVHMLI 1666
Db 1602 VVRVAGRLSEKMGKGPRAQOLRLAVCDLRLTNVAGVVDVYRXYGSPGLVHMLI 1661
QY 1667 GMLQTIADGKAHFTETIKPVLDTNLSIQRYE 1698
Db 1662 GMLQTIADGKAHFTETIKPVLDTNLSIQRYE 1693

RESULT 9
Q81862 PRELIMINARY: PRT: 1693 AA.
AC Q81862:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE ORF 1.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.C., Bradley D.W., Fry K.E.,
RA Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL Virology 185:120-131(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92261377; PubMed=1584074;
RA Uchida T., Suzuki K., Hayashi N., Iida F., Hara T., Oo S.S.,
RA Wang C.K., Shikata T., Ichikawa M., Rikihisa T., Mizuno K., Wain K.M.;
RT "Hepatitis E virus: cDNA cloning and expression.";
RL Microbiol. Immunol. 36:67-79(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92335008; PubMed=1630924;
RA Aye T.T., Uchida T., Ma X.Z., Iida F., Shikata T., Zhuang H.,
RA Wain K.M.;
RT "Complete nucleotide sequence of a hepatitis E virus isolated from the
RT Xinjiang epidemic (1988-1988) of China.";
RL Nucleic Acids Res. 20:3512-3512(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
RA Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271462; PubMed=1589964;
RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
RA Platak M., Feldman R.A., Yun K.Y., Purdy M.A., Et AL.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site.";
RL Virus Genes 6:173-185(1992).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=93348763; PubMed=8346669;
RA Bi S.U., Purdy M.A., McCausland K.A., Margolis H.S., Bradley D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL Virus Res. 28:233-247(1993).
DR EMBL: L08816; AAA03189.1;
DR INTERPRO: IPR000606;
DR INTERPRO: IPR002588;
DR INTERPRO: IPR002589;
DR PFAM: PF01443; Viral_helicase1;
DR PFAM: PF01660; Vmehylitransf. 1;
DR PFAM: PF01661; DUF27; 1;
SQ SEQUENCE 1693 AA; 185052 MW; D3363602435B8F5 CRC64;

Query Match 82.2%; Score 7412; DB 14; Length 1693;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1395; Conservative 106; Mismatches 174; Indels 32; Gaps 7;

QY 1 PGTTTAEQALAAANSALANAVVVRPPLSRVOTELINLMOPQOLVFRPVLLNHPQR 60
Db 10 PGTTTAEQALAAANSALANAVVVRPPLSRVOTELINLMOPQOLVFRPVLLNHPQR 69
QY 61 VIHNELEOYCARAGRCLEVGAPRSTINDPNVLRHCFRLRVGDRVORWYSAPRGPAN 120
Db 70 VIHNELEOYCARAGRCLEVGAPRSTINDPNVLRHCFRLRVGDRVORWYSAPRGPAN 129
QY 121 CRRSALRGLEPPADRTYCFDGSRCFAAETGVALYSLHDLMPADVAEAMARHGXTRLVAA 180
Db 130 CRRSALRGLEPPADRTYCFDGSRCFAAETGVALYSLHDLMPADVAEAMARHGXTRLVAA 189
QY 181 LHLPEVLLPRTGYHTHTSYLLIHGDRAVYTYEEDTSAGYVNDVSLRAMVTRTKIYGDH 240
Db 190 LHLPEVLLPRTGYHTHTSYLLIHGDRAVYTYEEDTSAGYVNDVSLRAMVTRTKIYGDH 249
QY 241 PLVIERRAICGCVLLLTAAPESPMPYVPYRSTEVYVRSIGPGSPSLFPSASCT 300
Db 250 PLVIERRAICGCVLLLTAAPESPMPYVPYRSTEVYVRSIGPGSPSLFPSASCT 309
QY 301 STFAVAVHIDRLMLFGATLDQAFCCSRMLTYLRGISYKVTVGALVANAGNASEDAL 360
Db 310 STFAVAVHIDRLMLFGATLDQAFCCSRMLTYLRGISYKVTVGALVANAGNASEDAL 369
QY 361 TAYTAAVLTICHOYRLTQTAISGMRRLGVENAKITRLIYSLFEKSGDYTPGROL 420
Db 370 TAYTAAVLTICHOYRLTQTAISGMRRLGVENAKITRLIYSLFEKSGDYTPGROL 429
QY 421 FYAOCRRMLSGFHLDRVAVLFDESVPCCRTPLKVAAGKCCFMRMLGOCCTFLPAE 480
Db 430 FYAOCRRMLSGFHLDRVAVLFDESVPCCRTPLKVAAGKCCFMRMLGOCCTFLPAE 489

```

QY 481 GLVGDHGHNEAVEGEVNDPAPRALHDVSGTYAVHGHOLEALYKRLNVPQDIAARASRLT 540
DB 490 GAVGOGHNEVEESDVPASALSDISGVYVPGTALQPLDYQALDPAETIYARRGRILT 549
QY 541 AVELVASPDRECFRTVLGNKFTFTTVVDGALHEANGPEQVYLSFDPASRQSGAGSHSLT 600
DB 550 AIVKVSQVNDGRIDCETILGNKFTFTSFDVGAVLETNGPRNHLSPDASSTYMAAGFSILT 609
QY 601 YELTPAGLQVRISSNGLDCTATFPPGAPSAAPGEVAAFCALRYRNRFQHRSLTGLGM 660
DB 610 YASAAGLEVRVYVAGLDHRAVFAVGPSPRSAPGEVTAFCSALYRREDAORSLTGNFM 669
QY 661 LAPEGLLGTFPPSPGHIMESANPFCGECTLYTRTMS-TSGSSDSP-----PEAANA 714
DB 670 FHPEGLGFAPSPGHIMESANPFCGECTLYTRTMSVADVSPAOPDLGFTSEPSIPS 729
QY 715 MAATGLPSTPVSADIWLPPESEFQVDAAPV-PPAPDPAQLP-GPVVLPPEPPPVH 772
DB 730 RAAT-----LTP-----AAPLPAPADPSPTPSAPARGEPAPGATAR 766
QY 773 KPSIP-PPSRNRLLYTPPDGAKVYAGSLFESDCOMLVNASNPGARGGGLCHAFYQRP 831
DB 767 APAITHQAARHRRLLFTYPDGSKVAFAGSLFESTCTWLVNASNVDRHGGGLCHAFYQRP 826
QY 832 EAFYTFEFTMRGLAAYTLTPRIIHAVAPDYRVQNPRLAAYRETSGRTAAYPL 891
DB 827 ASFDASAFVNRGGAAYTLTPRIIHAVAPDYRLHNPMLEAAVRETSGRTAAYPL 886
QY 892 GSGIYQVPSLSFDAMERHRRPDELTYLEPAPNPEAKPPQVYLTIEDARTANL 951
DB 887 GCGIYQVPSLSFDAMERHRRPDELTYLEPAPNPEAKPPQVYLTIEDARTANL 946
QY 952 EIDAAETEVGRACAGTISPGIVHYQFTAGVPGSGKSRSIQOGDVVDVYVPTRELNSMR 1011
DB 947 ELDSATDVGRACAGRVTFGVVQYQFTAGVPGSGKSRSITQADVVDVYVPTRELNSMR 1006
QY 1012 KGFAPFTPTTARVITGRVYVDEAPSLPHLLHMQASVYHLLGDPNOIPADIEHA 1071
DB 1007 KGFAPFTPTTARVITGRVYVDEAPSLPHLLHMQASVYHLLGDPNOIPADIEHA 1066
QY 1072 GLVPAIRPLAPTSMMXVYHRCPADVCYLIRGAYPKIOTSTRVLSLPMNEPAIGOKLV 1131
DB 1067 GLVPAIRPLAPTSMMXVYHRCPADVCYLIRGAYPKIOTSTRVLSLPMNEPAIGOKLV 1126
QY 1132 TQAAKAMPATVTHEAOGATFETTTIATADARGLIOSSRAHAIALTRHTEKCYILDA 1191
DB 1127 TQAAKAMPATVTHEAOGATFETTTIATADARGLIOSSRAHAIALTRHTEKCYILDA 1186
QY 1192 PGLREVGISDVIYVNNFLAGEVGNHRPSVTPRCNPDONLCTLOAPPSCQISAYHOLA 1251
DB 1187 PGLREVGISDVIYVNNFLAGEVGNHRPSVTPRCNPDONLCTLOAPPSCQISAYHOLA 1246
QY 1252 BELGRBPVAVANLPPCPLEEGLLYMPDELTVSDSVLFEELDIYHCHMAAPSOQKAVL 1311
DB 1247 BELGRBPVAVANLPPCPLEEGLLYMPDELTVSDSVLFEELDIYHCHMAAPSOQKAVL 1306
QY 1312 STLVGRGRTKLYEAHSDVRESLARFPTITGPVQATTCELYELVEAWEKQODSAYL 1371
DB 1307 STLVGRGRTKLYEAHSDVRESLARFPTITGPVQATTCELYELVEAWEKQODSAYL 1366
QY 1372 ELDDCNRDVSRITTFPOKCNKFTTGETTAHGVGOGISAMSKTFCALFGPWFARAIKAIL 1431
DB 1367 ELDDCNRDVSRITTFPOKCNKFTTGETTAHGVGOGISAMSKTFCALFGPWFARAIKAIL 1426
QY 1432 ALLPNIIEVGDAVESVFAAASGASGVFENDSEPTSTONNFTLGIECVMECCGMP 1491
DB 1427 ALLPNIIEVGDAVESVFAAASGASGVFENDSEPTSTONNFTLGIECVMECCGMP 1486
QY 1492 QMLIRLYHLVRSAMLIQAKESILKGFWKHSGEPGLTLNNTWYNNMAIIAHCEYFRDFRVA 1551
DB 1487 QMLIRLYHLVRSAMLIQAKESILKGFWKHSGEPGLTLNNTWYNNMAIIAHCEYFRDFRVA 1546

QY 1552 AFKGDQSVYLCSDPYROSNNAAALIAAGGLIKVYDPRIGLXAGVYVAPGLGTLPDVRRA 1611
DB 1547 AFKGDQSVYLCSEYROSPPAAVYLIACGLKLRKDFPRIGLXAGVYVAPGLGTLPDVRRA 1606
QY 1612 GRLEKNNMGPEPERAQLAVCDPLRGLTNVAQVQVVDVSVRYVYSGPLVHNLICMLQT 1671
DB 1607 GRLEKNNMGPEPERAQLAVCDPLRGLTNVAQVQVVDVSVRYVYSGPLVHNLICMLQT 1666
QY 1672 IADKAHFTETIKPVIDLTNSITIORVE 1698
DB 1667 VADKAHFTESYKPVLDLTNSITICRVE 1693
RESULT 10
Q9WC28 PRELIMINARY; PRT: 1693 AA.
AC Q9WC28;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE NON STRUCTURAL PROTEIN.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxId=12461;
RN [1]
RP SEQUENCE FROM N.A.
RA Ansari I.H., Nanda S.K., Durgapal H., Jameel S., Panda S.K.;
RT "Translational analysis of complete HEV genome.";
RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF076239; AAC27934.2; -
DR INTERPRO: IPR000606; -
DR INTERPRO: IPR002388; -
DR INTERPRO: IPR002589; -
DR PFM: PFM01443; Viral_helicase1; 1.
DR PFM: PFM01660; Umet_hyltransf; 1.
DR PFM: PFM01661; DMP27; 1.
SQ SEQUENCE 1693 AA; 185578 MW; FB37B87BF6BB501 CRC64;
Query Match 81.4%; Score 7333; DB 14; Length 1693;
Best local Similarity 79.8%; Pred. No. 0;
Matches 1375; Conservative 120; Mismatches 163; Indels 66; Gaps 6;
QY 1 PGITTAIEQAAALAAANSALANAVVBPFLSRVQTEILINLMOPROLVFRVLMNHPOR 60
DB 10 PGITTAIEQAAALATANSALANAVVBPFLSHQIEILINLMOPROLVFRVFMWNPOR 69
QY 61 VHNLEIQCRAAGACLEVGARPSINDPNVHLHRCFLRPVGRDQVQRTYSAPTGRPAAN 120
DB 70 VHNLEIQCRAAGACLEVGARPSINDPNVHLHRCFLRPVGRDQVQRTYSAPTGRPAAN 129
QY 121 CRSSALRGLPADRTYCFDGFSCAFPAETVALYSILHDLMPADVAEAMARHGXTRLYA 180
DB 130 CRSSALRGLPADRTYCFDGFSCAFPAETVALYSILHDLMPADVAEAMARHGXTRLYA 189
QY 181 LHLPEVLLPFGTYHTTSYLLIHDGRVAVTYEGDTSAGYNHDSILRAMIRTKIVGDH 240
DB 190 LHLPEVLLPFGTYHTTSYLLIHDGRVAVTYEGDTSAGYNHDSILRAMIRTKIVGDH 249
QY 241 PLVIERVARAIGCFVLLTAAPRSPMPYVPRYSREYVYRSIFGSGSPLFSPASCSTK 300
DB 250 PLVIERVARAIGCFVLLTAAPRSPMPYVPRYSREYVYRSIFGSGSPLFSPASCSTK 309
QY 301 STFHAVPVIHMDLMLFGATLDDQACCSRLMTYLGISYKVYVGLVLANEGNASSEDAL 360
DB 310 STFHAVPVIHMDLMLFGATLDDQACCSRLMTYLGISYKVYVGLVLANEGNASSEDAL 369
QY 361 TAVITAAVYLTICHQRYLQVLSKGMRLGVEHAQKFTIRLYSMLFEKSGRDYIPGRQLO 420
DB 370 TAVITAAVYLTICHQRYLQVLSKGMRLGVEHAQKFTIRLYSMLFEKSGRDYIPGRQLO 429
QY 421 FTAQCRRMLASGHHLDPRVLVFDSEVPCRCRFTLKLKAVACKFCFCFMWLOQECCTLEPAE 480
DB 421 FTAQCRRMLASGHHLDPRVLVFDSEVPCRCRFTLKLKAVACKFCFCFMWLOQECCTLEPAE 480


```

Db 430 FYACRRRLSGFHLDPVLYFDESAPCHCTVIRKALSKFCCKMKGQECTCLQPAE 489
QY 481 GLVGDHGDHNEAEGSEVDPAEPAPHLIDVSGYVAHGHOLEALRYALANVODIAAASBLT 540
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 GVVGDQGDHNSYEGSDVDPASIASDISGSIYVGTALQPLQYQALDDELVAACHLT 549
QY 541 ATVELVASPDLRECTVYCNKTFRTTYVDGAHLEANGPEQYVLSDSASROSKAGCSHSLT 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 550 ATVKASQYVDGIDCTELLGNKTFRTSFVDGAVLETNGBERHNLSPDASQSMAGPFSILT 609
QY 601 YELPAGLQVTRISSNGLDCTATEPPGAPSAAPGEVAAGFSAALYVNEFTQSHSLTGLM 660
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 610 YASASAGLEVYVAGAGLDHRAIFAPGVSPRNSPGEVTLAFCSALIKFNKEAQSHTLGLM 669
QY 661 LHPGLGLGIFPPFSPGHIMESANPECGEGTLTYRTWS-----TSGFSSDFSPPE 709
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 670 FHPGLGLGIFAPFSPGHIMESAKPFCGEGSTLYTRTWSVDVASSPTRDLFPMSE--PP- 726
QY 710 AAAPAMATPCLPHSTPVPVSDIWLPPPESEFOVDAAPVAPAPDAGLPGVYLTTPPPE 769
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 727 --TFSRAATPPL-----AAELP-----LAPDPS 749
QY 770 PVHPSIIPP-----SRNRLLTYTPDCAKYVAGSLFESDCDWLVNANP 814
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 750 PSSAPALDEPASATSGVPALTHTQAKRRLFTYPDGSKYFAGSLSESTLWLNASN 809
QY 815 GHRGGGGLCHAFYORPEAFYPTFEIMREGLAAYTLPRPIIHAVDPDYVEONPKRL 874
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 810 DHCGGGGLCHAFYORPEAFYPTFEIMREGLAAYTLPRPIIHAVDPDYVEONPKRL 869
QY 875 AYRTCSRGTAAPYPLGSGIYQVPSISFPAWERNHNPGEGLITTEPAANWFEANPAQ 934
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 870 AYRTCSRGTAAPYPLGSGIYQVPSISFPAWERNHNPGEGLITTEPAANWFEANPAQ 929
QY 935 PVLITTEDTARTANLATEIDATEVGRACACCTISPGIVHOFAGVGSKSSIOGD 994
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 930 PVLITTEDTARTANLATEIDATEVGRACACCTISPGIVHOFAGVGSKSSIOGD 989
QY 995 VDVVVVPTRELNSWRRRGFAAFTPHTAARVYIGRRVVIDEAPSLPHLLHMQRASSV 1054
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 990 VDVVVVPTRELNSWRRRGFAAFTPHTAARVYIGRRVVIDEAPSLPHLLHMQRASSV 1049
QY 1055 HILGDPNPIAIDFEBHGLVAIRPELAPTSWXYTHRCPPDVCCELLRGAVPKIQTTSRV 1114
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1050 HILGDPNPIAIDFEBHGLVAIRPELAPTSWXYTHRCPPDVCCELLRGAVPKIQTTSRV 1109
QY 1115 LRSLEWNEPAIGOKLYVTOAKKANPGAITVEHOGAFTETITIAADAGLIQSSRAH 1174
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1110 LRSLEWNEPAIGOKLYVTOAKKANPGAITVEHOGAFTETITIAADAGLIQSSRAH 1169
QY 1175 AIVALTHTTEKCVILDAAPGLLREVGISDVIYVNEFLAGGEVGHRRPSVYPRGNDONLGT 1234
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1170 AIVALTHTTEKCVILDAAPGLLREVGISDVIYVNEFLAGGEVGHRRPSVYPRGNDONLGT 1229
QY 1235 LQAPPPSCQISAHYQALAEELGHRAPAAVYPRCPPELQGLIYMPQELITYSDSVLEVLT 1294
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1230 LQAPPPSCQISAHYQALAEELGHRAPAAVYPRCPPELQGLIYMPQELITYSDSVLEVLT 1289
QY 1295 DIYCHRAAASORKAAYSTLVGRGRKTYEAHSDVRESLAFPTIGVQATTELY 1354
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1290 DIYCHRAAASORKAAYSTLVGRGRKTYEAHSDVRESLAFPTIGVQATTELY 1349
QY 1355 ELVEAMVEKQDSAVLELDCNRDVSRIFFQKCKNFTTGETIAGKVGOGISANSKT 1414
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1350 ELVEAMVEKQDSAVLELDCNRDVSRIFFQKCKNFTTGETIAGKVGOGISANSKT 1409
QY 1415 FCALFGWFAIEKAILLPPNIFYDAYEESVFAAASGAGSCWVFENDFSEEDSTON 1474
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1410 FCALFGWFAIEKAILLPPNIFYDAYEESVFAAASGAGSCWVFENDFSEEDSTON 1469
QY 1475 NFGSLGECVVMESGMPQWLIRLYHLVRSAMIIQAPKESLKGPKKKSGEPGILLMNTV 1534
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1470 NFGSLGECVVMESGMPQWLIRLYHLVRSAMIIQAPKESLKGPKKKSGEPGILLMNTV 1529

```

```

QY 1535 NMAIACHYERDFEVAFAFKDDSVYLCSDYRQSHNAALAGGLAKLVDPYRIGLYAG 1594
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1530 NMAVITHCYDRDLQVAFKDDSDIVLCSEYRQSGAVALIAGGLAKLVDPYRIRLYAG 1589
QY 1595 VVAAGLCTLDVVRFAGLSEKMKWGPGERAEOLRLAVCDFLGLTNVAQVCDVVS 1654
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1590 VVAAGLCTLDVVRFAGLSEKMKWGPGERAEOLRLAVCDFLGLTNVAQVCDVVS 1649
QY 1655 YGVSGLVHNLIGMQLTADGKAHFTETIRKPVLLTNSIIORVE 1698
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1650 YGVSGLVHNLIGMQLTADGKAHFTETIRKPVLLTNSIIORVE 1693

RESULT 11
ID 039221 PRELIMINARY; PRT: 1693 AA.
AC 039221;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDEIN.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=HYDERABAD, INDIA;
RA Asari I.H., Nanda S.K., Durapal H., Jameel S., Panda S.K.;
RT "Eukaryotic expression of nonstructural protein (ORF1) of HEV: absence
RT of any processing."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028091; AAB82002.2;
DR INTERPRO: IPR006066;
DR INTERPRO: IPR002588;
DR INTERPRO: IPR002589;
DR PFM: PFM1443; Viral helicase; 1.
DR PFM: PFM1660; Vmehylitransf. 1.
DR PFM: PFM1661; DUF27; 1.
KW Polypeptidein.
SQ
SEQUENCE 1693 AA: 185562 MW: 7AFED003532E12BC3 CRC64:

Query Match 81.3%; Score 7329; DB 14; Length 1693;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1374; Conservative 121; Mismatches 163; Indels 66; Gaps 6;
QY 1 PGITTAIEQAAALAAANSALANAVVVRPPLSRVQTEIILNMQPQOLVFRPEVIANHPIDR 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 PGITTAIEQAAALAAANSALANAVVVRPPLSRVQTEIILNMQPQOLVFRPEVIANHPIDR 69
QY 61 VINHELOYCARARGCLTEGAHPRSTINDPNVYHRCFLRPVGDVQRYASAPRGPAAN 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 70 VINHELOYCARARGCLTEGAHPRSTINDPNVYHRCFLRPVGDVQRYASAPRGPAAN 129
QY 121 CRSSALRGLPADRTYCFDFSSRCFAETGVALYSLHDLMPADVADAMAHGXTRLYAA 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 CRSSALRGLPADRTYCFDFSSRCFAETGVALYSLHDLMPADVADAMAHGXTRLYAA 189
QY 181 LHPPEVLPRGYHTTSYLLIHDGDAVVYBGSYSAGYNHDSILRAMIRTKIYGDH 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 LHPPEVLPRGYHTTSYLLIHDGDAVVYBGSYSAGYNHDSILRAMIRTKIYGDH 249
QY 241 PLVIERRAIGCHEVILLTAPAPSPMPVYPRSTVYVRSIFGPGSGSLSPSACSTK 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 PLVIERRAIGCHEVILLTAPAPSPMPVYPRSTVYVRSIFGPGSGSLSPSACSTK 309
QY 301 STFHAHVHIMDRMLFGATLDDOAFCCSRMLTYLRGISYKVYVGAIVANEGNNASEDAL 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 STFHAHVHIMDRMLFGATLDDOAFCCSRMLTYLRGISYKVYVGAIVANEGNNASEDAL 369
QY 361 TAXITAAVLTICQORYLTCQAIKGMRLGVEHAQKFTTLYLWLEKSGRDYIPGROLQ 420
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

Db 370 TAVITAAYLTICHOVRLTFOAISKIRLEREHQKFIITRLXSWLFESKSGRDYIPGRQLE 429
Qy 421 FTAQCRRLISAFHLDPRVLVDESVPCRCRTFLKLVAGKFCCEPMRLQOECTCELEPME 480
Db 430 FTAQCRRLISAFHLDPRVLVDESVPCRCRTFLKLVAGKFCCEPMRLQOECTCELEPME 489
Qy 481 GLVGHGHNEVYEGSEVDPAEPALDVSGTVAVGHGHOEALYRALNVPDIAAARSRLT 540
Db 490 GVVGGQGHNEVYEGSEVDPAEPALDVSGTVAVGHGHOEALYRALNVPDIAAARSRLT 549
Qy 541 ATVELVAPSDRELECTVLGNKTFRTTVVDGAHLLEANGPEOYVLSFDASROSAGASHSLT 600
Db 550 ATVKVSOYQDRIDCELTGLNKTFRFSYVDGAVLETLNGPRNHLSPDASOSTMAAGFSILT 609
Qy 601 YELPAGLQVRISSNGLDCTATPPGAGSAAPGEVAACSAALRYNRFTQHSILTGILM 660
Db 610 YAAASAGLEVRYVAGLDHRAIFAGVSPRSNPGEVTAFCASALYFRNREAOHRSILTNIM 669
Qy 661 LHPEGLLIFPPSPGHITMESANPFCGEGTLTTRTWS-----TSGFSSDPSPPE 709
Db 670 FHPEGLLIGLFAFSPGHWESAKPFCGEGTLTTRTWS-----TSGFSSDPSPPE 726
Qy 710 AAAPMAATPGDLPSTPVDIWLPPSEEFQVDAAPVPAAPDAGLPGVVLTPPPPP 769
Db 727 --TPSRATPPL-----AAAPLP-----LAPDSP 749
Qy 770 EVHKPSIPP-----SRNRLLYTPDGAKVYAGSLFESDCMLVANSP 814
Db 750 PSSAPALDEPASAATSGVPAITHQTAHRRLLETFYPDGSKVPAGSLFESTCMLVANSP 809
Qy 815 GHRPGGGLCHAYORPREFAFTEFIMREGLAAYTLTPRIIHAAPDVRYQNPRLRLA 874
Db 810 DHCPGGGLCHAYORPREFAFTEFIMREGLAAYTLTPRIIHAAPDVRYQNPRLRLA 869
Qy 875 AYRETCRSRGTAAYPLDGSIGYQVPSLSFDAMERNHRGDELYLEPAANFEAKPPAQ 934
Db 870 AYRETCRSRGTAAYPLDGSIGYQVPSLSFDAMERNHRGDELYLEPAANFEAKPPAQ 929
Qy 935 PVLTTTETDARANTALEIDAATENVGRACAGCTISPGIYHYOFTAGVPGSGKSRSTIQGD 994
Db 930 PVLTTTETDARANTALEIDAATENVGRACAGCTISPGIYHYOFTAGVPGSGKSRSTIQGD 989
Qy 995 VVVVVVPRTELNSMRGFAFTPTARVITGRVVVDEAPSLRPHLLHMQAASSV 1054
Db 990 VVVVVVPRTELNSMRGFAFTPTARVITGRVVVDEAPSLRPHLLHMQAASSV 1049
Qy 1055 HLLGDPNQIPADIDFEHAGLVPARPELAPTSMWXYTHRCPADVCELIRGAYPKIQTTSRV 1114
Db 1050 HLLGDPNQIPADIDFEHAGLVPARPELAPTSMWXYTHRCPADVCELIRGAYPKIQTTSRV 1109
Qy 1115 LRSLEWNEPAIQOKLVYTOAANKAANGAIVHEAOGATETETIATADARGLIQSSRAH 1174
Db 1110 LRSLEWNEPAIQOKLVYTOAANKAANGAIVHEAOGATETETIATADARGLIQSSRAH 1169
Qy 1175 ALVALTRHTEKVIIDAPGLREVGISDYIVNNEFLAGGEIGHQRSVIPRGPDDNLTG 1234
Db 1170 ALVALTRHTEKVIIDAPGLREVGISDYIVNNEFLAGGEIGHQRSVIPRGPDDNLTG 1229
Qy 1235 LOAFPPSCQISAYHOLAEELGHRPAVAVLPPCPLEGLLMPOLTVSDSVLEFLT 1294
Db 1230 LOAFPPSCQISAYHOLAEELGHRPAVAVLPPCPLEGLLMPOLTVSDSVLEFLT 1289
Qy 1295 DIVHCRMAAPSORKAVLSTLVGRYGRRTKLYEASHDSVRESLARPTPTGPVOATTCELY 1354
Db 1290 DIVHCRMAAPSORKAVLSTLVGRYGRRTKLYEASHDSVRESLARPTPTGPVOATTCELY 1349
Qy 1355 ELVEAMVEKGGGSAVLELDLNRDYSRTTFPOKXCNKFTTGETTAHGAVGQGISMSMT 1414
Db 1350 ELVEAMVEKGGGSAVLELDLNRDYSRTTFPOKXCNKFTTGETTAHGAVGQGISMSMT 1409
Qy 1415 FCALFGPFRATELEKILLAPNIFYGAYEBSVFAAASGAGSCAVFENDESEPDSTON 1474
Db 1410 FCALFGPFRATELEKILLAPNIFYGAYEBSVFAAASGAGSCAVFENDESEPDSTON 1469

```

```

Qy 1475 NFSLIGECVYMECCGMQWILIRLYHRSAMTILQAPKESLKGPRKHSBEPGLTMMTV 1534
Db 1470 NFSLIGECVYMECCGMQWILIRLYHRSAMTILQAPKESLKGPRKHSBEPGLTMMTV 1529
Qy 1535 NMAIITHCEYFDFRYAAAKGDDSVYLCSDYROSMAALLIGCCGKLKVDYPIGLYAG 1594
Db 1530 NMAIITHCEYFDFRYAAAKGDDSVYLCSDYROSMAALLIGCCGKLKVDYPIGLYAG 1589
Qy 1595 VVAPGLGLPDPVPRAGRLSEKNNMGPRGPERAEQRLAVCDPLRGLTNYAQCVDVYSRV 1654
Db 1590 VVAPGLGLPDPVPRAGRLSEKNNMGPRGPERAEQRLAVCDPLRGLTNYAQCVDVYSRV 1649
Qy 1655 YGVSPGLVHNLIGMLQAVDAGKHFESYKPYVLDLNSILCRVE 1698
Db 1650 YGVSPGLVHNLIGMLQAVDAGKHFESYKPYVLDLNSILCRVE 1693

RESULT 12
Q81344 PRELIMINARY; PRT; 1693 AA.
AC Q81344;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE HEPATITIS E VIRUS COMPLETE GENOME.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
RN NCBI_TaxID=12461;
RX MEDLINE=93348763; PubMed=8346669;
RA Bi S.L., Purdy M.A., McCausland K.A., Margolis H.S., Bradley D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China."
RL Virus Res. 28:233-247(1993).
DR EMBL: M94177; AAA96139.1; -.
DR INTERPRO: IPR000606; -.
DR INTERPRO: IPR002588; -.
DR INTERPRO: IPR002589; -.
DR PFAM: PF01443; Viral_helicase1.
DR PFAM: PF01660; Vmethyltransf. 1.
DR PFAM: PF01661; DUF277.1.
SQ SEQUENCE 1693 AA; 185271 MW; E330BF128EDB9C6F CRC64;

Query Match 81.0%; Score 7304; DB 14; Length 1693;
Best Local Similarity 80.8%; Pred. No. 0;
Matches 1380; Conservative 103; Mismatches 192; Indels 32; Gaps 7;

Qy 1 PGTTAIEQAALAAANSALANAVVPRFLSRVQTEILINIMOPROLVFREVLMNHPOR 60
Db 10 PGTTAIEQAALAAANSALANAVVPRFLSHQIEILINIMQROQLVFEVEMNHPOR 69
Qy 61 VHNELLEYCRARAGCLEYGAPHSINDPNVYLHRCFLRPVGRDYQKWTSAATRGPAN 120
Db 70 VHNELLEYCRARAGCLEYGAPHSINDPNVYLHRCFLRPVGRDYQKWTSAATRGPAN 129
Qy 121 CRSSALRGLPAPARVYCPGFGFSCAFPAETGVALYSIHDMPADVAEAMRHGXTRLYAA 180
Db 130 CRSSALRGLPAPARVYCPGFGFSCAFPAETGVALYSIHDMPADVAEAMRHGXTRLYAA 189
Qy 181 LHLPEVLLPPGTYHTTSTYLLIHDGDAVVTYBGDTSAGYNHDSILRAMIRTKIYGDH 240
Db 190 LHLPEVLLPPGTYHTTSTYLLIHDGDAVVTYBGDTSAGYNHDSILRAMIRTKIYGDH 249
Qy 241 PLVIEEVRALGCFVULLTAAPESPMPYVPRPRSTEVYKRSITFGGSGSLPSSACSTK 300
Db 250 PLVIEEVRALGCFVULLTAAPESPMPYVPRPRSTEVYKRSITFGGSGSLPSSACSTK 309
Qy 301 STPHAVPVHMDLMLFGATLDDQAFCCSRLMYLIGISYKVVGVGLVANEGNASPDAL 360
Db 301 STPHAVPVHMDLMLFGATLDDQAFCCSRLMYLIGISYKVVGVGLVANEGNASPDAL 360

```

D	b	310	RPHNAVPHAMIDRLMLGCAITLDDQAPCCSNLMYTLNGISVKKYVGLVLANBGMNMSVVAL	369
Q	y	361	TAXITAAVLTICHOYRTLTOAISKMRRLCEVHAOKFTIRLSWFEKSGRDYIPGRLO	420
D	b	370	TAVITAAVLTICHOYRLTQTAISKMRRLREHAOKFTIRLSWFEKSGRDYIPGRLO	429
Q	y	421	FYACRRRLASGPHLDPRVLVPEDESVCRCRTEFLKKVYAKPCCPMWLGODETCLEPAE	480
D	b	430	FYACRRRLASGPHLDPRVLVPEDESAPCHCRTAIRKAVSFFCCPMWLGODETCLOAPE	489
Q	y	481	GLVGDHGHNEAYGSEVDAEPABHLDDSGTAVVHCHOLEALYRALNPODIAAASRLT	540
D	b	490	GAVDGDGHNEAYGSDVDAEASLSDISSYVPEPTALQPLQALDPAELVANAAGLIT	549
Q	y	541	ATVELVASPDRECRYLGNKTRRTTYVDCAHLEANGPEQOYVLSFDSQSGAGSHLT	600
D	b	550	ATVAVSOVDKIDDETLGLNKTERTSFTVGCAYVLETNGPERHNLSPDASQSTMAAPFSIT	609
Q	y	601	YELPAGIAYRISNGIDCATFPPGAGPAAACEVAACSAALYRNRTQHSITLGLGM	660
D	b	610	YASASAGLEVRYAAGADHRAVAPGVSPPASAGCEYATCASALYRNRAQRLSLTGFW	669
Q	y	661	LHPGELGIFPPEPGHIMESANPFCGEGLYRTWS-TSGSEDPSP-----PEAAPA	714
D	b	670	FHPGELGIFPPEPGHIMESANPFCGEGLYRTWSSEVDANASSPAQPOLGISEPSTPS	729
Q	y	715	MAATPGLPHSTPPVSDIWLVPPESEEFQVDAAPV-PPADPAGLP-GRVVLTPPPPVH	772
D	b	730	RAAT-----LTP-----AAPLPAPDPSPTSPAPARGEPAGATAR	766
Q	y	773	KPSLP-PESNRRLTYPPDGAAYVAGSLPESQDMLVNASNPGRRPGGLCHATYQFR	831
D	b	767	APATTHOARRRLLEFYPGCSKVPAGSLFESCTYLVNASVNDHPRGGGLCHAYQRP	826
Q	y	832	EAFPTFERIMEGILAAATLTPRPLIHAVAPDYVEQDNPRLAEARETCSRGTAAAYPL	891
D	b	827	ASFPASCFVMDGAAATLTLPRLIHAVAPDYLENNPMLAEARETCSRGTAAAYPL	886
Q	y	892	GSGIYQVPSLSFDAMERNHRPDELYLTERPAANMEEANKPAQVLTITTEDTARTANAL	951
D	b	887	GTGICYOVPISGFEPMERNHRPDELYLTERPAANMEEANKPAQVLTITTEDTARTANAL	946
Q	y	952	EIDATEVGRACAGCTISPEIYHIOFTAGYPGSGKRSITQCGDVVVYPTRELNSMR	1011
D	b	947	ELDSATVGRACACRYTPGVQYQFTAGYPGSGKRSITQADVVDVYPTRELNSMR	1006
Q	y	1012	RGFAFPHTPAARTITRRVYIDEAASLPRIHLLIMQASSHLLIGDNOIPIADEFHA	1071
D	b	1007	RGFAFPHTPAARTITGRRAVYIDEAASLPRIHLLIMQAAATVHLLIGDNOIPIADEFHA	1066
Q	y	1072	GLVPAIRPELAPTSMWVYVTHRCPADVCELRGAYPKIQTTSRYVLSLFENNEPAIGOKLVX	1131
D	b	1067	GLVPAIRPELAPTSMMVYVTHRCPADVCCELRGAYPKIQTTSRYVLSLFENNEPAIGOKLVX	1126
Q	y	1132	TOAKKANPGATVYHEAQGTTEETITATADARGLIOSSRAHAIVALTIRTEKCVILDA	1191
D	b	1127	TOAKKANPGSVYVHEAQGTTEETITATADARGLIOSSRAHAIVALTIRTEKCVILDA	1186
Q	y	1192	PGLLREGISDVYNNPEFLAGVGWGHRSVYPRGPDQNGICTQAFPPSOQISAVYHOLA	1251
D	b	1187	PGLLREGISDAILYNNPEFLAGVGELGHQORSVYPRGPDQNAVDTLLAAPPSCQISAFHOLA	1246
Q	y	1252	EELCHRAPAPAAVLPPELEQGLKAMPQELVYSDSVLFEELTDIVHOCMAAPSORKAVL	1311
D	b	1247	EELCHRAPAPAAVLPPELEQGLKAMPQELVYSDSVLFEELTDIVHOCMAAPSORKAVL	1306
Q	y	1312	STLVGRYGRFKLYEAASDVRESLARFTPTIGPVQATTCEDLYELVEAMVEKGODGSAYL	1371
D	b	1307	STLVGRYGRFKLYNASHSVPSDSLARFTPAIGHQVTTCELLYELVEAMVEKGODGSAYL	1366
Q	y	1372	ELDLCNDVSRITFPQKXCKFTTGETITAHGKVGCISAMSTFCALFGPMFRALEKITL	1431
D	b	1367	ELDLCNDVSRITFPQKCKFTTGETITAHGKVGCISAMSTFCALFGPMFRALEKITL	1426

[illegible]

RESULT	13		
Q069418			
ID	Q069418	PRELIMINARY:	PRT: 1693 AA.
AC	Q069418:		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)	
DE	ORF1, ORF2 & ORF3.		
OS	Hepatitis E virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage.		
OX	NCBI_TaxID=12461;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=INDIVIDUAL PATIENT INFECTED WITH HEV;		
RA	Von Brunn A., Seebach J., Thyagarajan S.P., Mohanavalli B., Menon T.,		
RA	Froesner G.;		
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: X99441; CA67802.1; -.		
DR	INTERPRO: IPR000606; -.		
DR	INTERPRO: IPR002588; -.		
DR	INTERPRO: IPR002589; -.		
DR	PFAM: PF01443; Viral_helicase1; 1.		
DR	PFAM: PF01660; Methyltransf. 1.		
DR	PFAM: PF01661; DUF27; 1.		
Q0	SEQUENCE 1693 AA; 185949 MW; 489B4ADBA5E7E529 CRC64;		

Query Match	81.0%;	Score 7301;	DB 14;	Length 1693;
Best Local Similarity	79.9%;	Pred. No. 0;		
Matches 1367;	Conservative 116;	Mismatches 188;	Indels 40;	Gaps 6;
QY	1	PGITTAIEQALAAANSALANAVVPRPELSRQTEIILINLMOPROLVREPEVLMNPIOR	60	
Db	10	PGITTVIEQALAAANSVLANAVVPRPLSHQIIEILINLMOPROLVRRPEVFMNHPIOR	69	
QY	61	VIHNELEOYCARAGCCLVEGAPRSINDNPVYLRCFELRPVGRDQRYTSAPTRGPAN	120	
Db	70	VIHNLELYCARASRCLEIGHAPRSINDNPVYARCFELRPVGRDQRYMHPATRGPDAN	129	
QY	121	CRRSALRGLPADRDRTYCFDGFSCAFAAETGVALYSLHDMLPADVAEAMRHGATRLTAA	180	
Db	130	CRRSALRGLPADRDRTYCFDGFSGCFNFPATGVALYSLHDMSPSDVAEAMFRHGCMRTLYAA	189	
QY	181	LHLRPEVLLPFGYTHHTSYLLIHGDRAVVYIEGTSAGYNHDSVILEAMIRTKTIVGDH	240	
Db	190	LHLRPEVLLPFGYRTASYLLIHGDRVVVYIEGTSAGYNHDSVILNLSWIMTKTIVTGDH	249	
QY	241	PLVIERVAICCHVLLITTAPEPSPMYVYPRSTEEVYVRSIFEGPGSPSLFSPCASCTK	300	
Db	250	PLVIERVAICCHVLLITTAPEPSPMYVYPRNSTEEVYVRSIFEGGPGPSLFLPSCSCTK	309	

QY	301	STFAVNVHIMDRIMLMEGATLDQACRSRIMTYLRGISYKVTQVATLVANGMNASDPL	360
Db	310	STFAVNAHIMDRIMLMEGATLDQACRSRIMTYLRGISYKVTQVATLVANGMNASDPL	369
QY	361	TAXITAAVLTICHOVYLRTOAISKGRRIYGVSHAQCFITRLYSWLFEEKSGRDYIPGRLO	420
Db	370	TAVITAAVLTICHOVYLRTOAISKGRRIYGVSHAQCFITRLYSWLFEEKSGRDYIPGRLO	429
QY	421	FYAOCRRWLSAGFHLDPRVLVEDESVPCRCRTFLKVKVAGFECCEMRWLGOCSTGLEBAE	480
Db	430	FYVOCRRWLSAGFHLDPRVLVEDESVPCRCRTFLKVKVAGFECCEMRWLGOCSTGLEBAE	489
QY	481	GVNDHGHOMEAESSEVDPAERAHLDVSGTVAHVGHOLEALRYLRANYPDOLIAARASLT	540
Db	490	GVNDHGHOMEAESSEVDPAERAHLDVSGTVAHVGHOLEALRYLRANYPDOLIAARASLT	549
QY	541	ATVELVASPDRECEPRVYLGNKTRFRTVYDGAHLEANGPEOYVLPDASROSAGASHLT	600
Db	550	ATVAVSVDORIDICEFILLGNKTRFRTFVYDGAHLEANGPEOYVLPDASROSAGASHLT	609
QY	601	YELTPAGLOVRISNGLDCTATFPFGGABSAARGEVAAFCSALYKYNFTQJHNSITGLGM	660
Db	610	YAVSADLEEVRYAAGLDHRAVAPFVSRSRPFGEVATFCSALYKYNFTQJHNSITGLGM	669
QY	661	LHPRGILGTIPRSPBGHIMESANPCGECTLYTRTWS-TSGSSDSP-----PEAARA	714
Db	670	FHPRGILGTIPRSPBGHIMESANPCGECTLYTRTWS-TSGSSDSP-----PEAARA	719
QY	715	MAAPRGPHSTPVPVDIWLPRPSEEFQYDAPV-PRAPDPAGLGPVYLTPRPPRPNK	773
Db	730	RAAPRPL-----AAPRPPADPSPPS-----APADPRASG	762
QY	774	PSLPP-----SRNRLLYTPYDGAKYAGSLFESDCOMLVASNPGRHPRGGGLCHAFY	827
Db	763	TTAGAPRITQHMRHRRLLFTYPDGSKVPAGSLFESTCTWPNASGVNDRHPRGGGLCHAFY	822
QY	828	QRPEAFATPREFIMRGLAATITLTPRPIIHAAVDYRVQNKRLLEAAYRETSARGTAA	887
Db	823	QRYPASDASAFYMRGGAATLTPRPIIHAAVDYRVQNKRLLEAAYRETSARGTAA	882
QY	948	NLALEIDAATEVGRACAGCTISPFIYHOFYTAGVPGSGKSRSIOGDVDVVPVPTRELIN	1007
Db	943	NLALEIDAATEVGRACAGCTISPFIYHOFYTAGVPGSGKSRSIOGDVDVVPVPTRELIN	1002
QY	1008	SWRRGRGAFTPTTAARVITGRRVYIDEAPSLRPHLLIHMRASSVHLLDGPNOIPAD	1067
Db	1003	AMRRGRGAFTPTTAARVITGRRVYIDEAPSLRPHLLIHMRASSVHLLDGPNOIPAD	1062
QY	1068	FEHAGVLPATRPDLAPTSMXHVYHRCRPAVCELTGAVKPIOTGSRVYLSLFWMPRAIG	1127
Db	1063	FEHAGVLPATRPDLAPTSMXHVYHRCRPAVCELTGAVKPIOTGSRVYLSLFWMPRAIG	1122
QY	1128	KLVYTOAKAANPGATITVEAOGATTEETTITATADARGLIOSSRAHIVALTTRHTEKV	1187
Db	1123	KLVYTOAKAANPGATITVEAOGATTEETTITATADARGLIOSSRAHIVALTTRHTEKV	1182
QY	1188	ILDAPEGLREVGISDYIVNNFFLAGGEVGHRRSVIPRGNDONIGTLOAFRPSQOISAY	1247
Db	1183	ILDAPEGLREVGISDYIVNNFFLAGGEVGHRRSVIPRGNDONIGTLOAFRPSQOISAY	1242
QY	1248	HQLEELGHPAPVAVANLPCRELTBOGLLYMOEQLTSDSVLYEFLTQIVICRMAAPSR	1307
Db	1243	HQLEELGHPAPVAVANLPCRELTBOGLLYMOEQLTSDSVLYEFLTQIVICRMAAPSR	1302
QY	1308	KAVYSTLVGRYGRRTKUYEAHSDVESLAREIPTGPQVATCELYELVAMAEKGODG	1367
Db	1303	KAVYSTLVGRYGRRTKUYEAHSDVESLAREIPTGPQVATCELYELVAMAEKGODG	1362
QY	1368	SAVLELDLCNRDYSRIITEFOCKXCNFTTGETTAHGKVGOGISAWSKTECALGFWFAIE	1427

D	b	1363	SAYLEHEMNDVSTLITFQDDCKNFTTGEIAHAKVGQGISASKTCTPLGPFRAIE	1422
O	y	1428	KEILLALPPNFPGAYEESYFAAASGAGSCGMENFSEFDSTONNFSLGLECVMAE	1487
D	b	1423	KAILALLPQGFGFYGDAPFDPLYFAAIVPAKRSWMEFNFSFSDSQNNFSLGLECAIMEE	1482
O	y	1488	CGMPOMLIRLYHLVASMIILQAPKESLEGFKKHSGEPGTLLMTVMMAIIAHCEFRD	1547
D	b	1483	CGMPOMLIRLYHLIRSAMILOAPKESLEGFKKHSGEGPTLLMTVMMAIYTHCYDERD	1542
O	y	1548	FRVAAFKDDSVLICSDVRORRNAALLIAGCGIKTKVDYRPIGLYAGVVAPGLGTEPDY	1607
D	b	1543	LQVAAPFKDDSVLICSESRQSPGAANVLLAGCGIKTKVDREPIGLYAGVVAPGLGALPDV	1602
O	y	1608	VRFAGRLSEKMMWGPEPEAEOLRLAVCDFGLGTJNVACVADVVSRYVGSFGLVHNLIIG	1667
D	b	1603	VRFAGRLSEKMMWGPEPEAEOLRLAVSDFELKLTINVAOCVADVVSRYVGSFGLVHNLIIG	1662
O	y	1668	MLQTLADGKAHFETIKFPVLDLTNSIIORVE	1698
D	b	1663	MLQAVADGKAHFETESVKRPVLDLTNSILCRVE	1693
R	E	RESULT	14	
I	D	081873	PRELIMINARY;	PRT; 727 AA.
A	C	081873;		
D	T	01-NOV-1996	(TREMBLrel. 01, Created)	
D	T	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
D	T	01-JUN-2000	(TREMBLrel. 14, Last annotation update)	
D	E	POLYPROTEIN		
D	E	(ENCODING NTP-BINDING PROTEIN AND RNA-DEPENDENT RNA POLYMERASE)		
D	E	(FRAGMENT).		
O	S	Hepatitis E virus.		
O	S	Vituses; ssRNA positive-strand viruses, no DNA stage.		
O	X	NCBI_TaxID=12461;		
R	N	[1]		
R	N	SEQUENCE FROM N.A.		
R	A	Tam A.W., Smith M.M., Kim J.P., Young L.M., Platak M., Feldman R.A.,		
R	A	Purdy M.A., Bradley D.W., Reyes G.R., Fry K.E.;		
R	L	Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).		
R	N	[2]		
R	P	SEQUENCE FROM N.A.		
R	X	MEDLINE=90193694; PubMed=2107574;		
R	A	Reyes G.R., Purdy M.A., Kim J.P., Luk K.C., Young L.M., Fry K.E.,		
R	A	Bradley D.;		
R	T	"Isolation of a cDNA from the virus responsible for enterically		
R	T	transmitted non-A, non-B hepatitis.";		
R	T	Science 247:1335-1339(1990).		
R	N	[3]		
R	P	SEQUENCE FROM N.A.		
R	X	MEDLINE=92271462; Pubmed=1589964;		
R	A	Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,		
R	A	Platak M., Feldman R.A., Yun K.Y., Purdy M.A., Et AL.;		
R	T	Hepatitis E virus (HEV): strain variation in the nonstructural gene		
R	T	region encoding consensus motifs for an RNA-dependent RNA polymerase		
R	T	and an ATP/GTP binding site.";		
R	L	Virus Genes 6:173-185(1992).		
R	L	EMBL; M32400; AAA03206.1; -.		
D	R	INTERPRO: IPR000606; -.		
D	R	PFAM: PF01443; Viral_helicase1; 1.		
D	R	Polyprotein; RNA-directed RNA polymerase.		
F	T	NON_TER	1	
S	O	SEQUENCE	727 AA; 79306 MW; 872631FC62318FA7 CRC64;	
Q	U	Query Match	38.2%; Score 3444; DB 14; Length 727;	
Q	U	Best local similarity	88.7%; Pred. No. 4.6e-246;	
Q	U	Matches 645; Conservative 40; Mismatches 42; Indels 0; Gaps 0;		
Q	U	1VHVQTAGVGCSSGRSIQQGGDVVVVPTRELNSNRRCGAATPHTAARTIGRRV	1031	

